New Labeling Strategies to Investigate the Folding of Large RNAs via FRET

DISSERTATION

zur

Erlangung der naturwissenschaftlichen Doktorwürde

(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Anita G. Schmitz

aus

Deutschland

Promotionskomitee

Prof. Dr. Roland K. O. Sigel (Vorsitz)

Prof. Dr. Gilles Gasser

Dr. Eva Freisinger

Zürich, 2014

Acknowledgments

I would like to thank:

My supervisor Prof. Dr. Roland K. O. Sigel for giving me the opportunity to do my PhD project in his group, for the challenging topic and the support. It was great to learn so many techniques and have the freedom to realize in my ideas.

The members of my PhD committee, Prof. Dr. Gilles Gasser and Dr. Eva Freisinger for helpful discussions. Especially Gilles for the very good collaboration and support with the PNA project.

The SF-Lab (Sigel-Freisinger group) for the good working atmosphere. It was always nice to get a smile or having a chat in the corridors or labs. Also for great lab events like countless BBQs, boat trips, parties, football championship watching, swimming, foxtrail, hiking, ski weekends, Sola and so on.

My subgroup leader Dr. Sofia Gallo for the introduction in the techniques of RNA, and for the continuous support. Also for all the administrative and organizational tasks. Dr. Silke Johannsen for encouraging me to stay positive and giving me another view on things. Thanks to my 2nd subgroup leaders Dr. Danny Kowerko and Dr. Richard Börner for all the discussions about FRET. Danny for amusing trips to Germany with long discussions about photophysics. Susann Zelger-Paulus for giving me so much feedback and great ideas for this project. And also for showing me how to be a good RNA biochemist, for the best T7 polymerase in the lab and the very fruitful collaboration including the PCR of the new constructs. Helena Guiset Miserachs for being such a good friend in and outside the lab and for motivating me. Dr. Miriam Skilandat for good discussions, feedback, and the nice time besides the lab. My lunchgroup 11:30 for the nice moments, having fun and giving new ideas and energy for the day (Silke, Susi, Helena, Danny, Richard, Miriam, David, Tamara, Jay, Maria, Maxi, Ramona). Mélodie Hadzic for the SIRA aka MASH software, for solving every error message within seconds and giving me the opportunity to give feedback and for all the support with the smFRET data evaluation. Erica Fiorini for the labeled 17/7 and making the radiolab a bit more bearable with her presence and the music. All the smFRET subgroup members Mélodie, Erica, Mokrane, Sebastian, Susi, Helena, Danny and Richard and the ChemBiol subgroup members Sofia, Anastasia Musiari, Dr. Michelle Schaffer and Dr. Pallavi Choudhary for the helpful discussions. Anastasia and Sofia for the pPC1 RNA and being great *btuB*-B_{12} mates. Dr. Miriam Steiner for leaving these wonderful labbooks which were so easy to follow. Dr. Daniela Donghi for the organization of the gel lab. Dr. Joe Nagel alias

Joachim Schnabl, Dr. Magdalena Rowińska-Żyrek, Dr. Maria Pechlaner, Jovana Jakovleska and many others mentioned above for the nice times in and outside the lab.

The past group members Dr. Maximilane Korth, Dr. Miquel Barceló Oliver and Dr. Lucia Cardo for a good start. Miquel for the support with the B_{12} project and synthesis of dye precursors. The students Sonja Giger and Adrian Kämpfer for their contributions to the dye syntheses. My dear friend Dania Marthaler for the continuous support during my PhD time and for being the best labmate. Dr. David Egloff for good discussions about chemistry and HPLC, teaching me Swiss language and traditions and being a good friend besides lab.

The Gasser Group for being my second group and all the nice time together. Especially Dr. Anna Leonidova and Dr. Philipp Anstätt for the introduction and help in PNA synthesis. Anna also for the MS-measurements and for trusting me with her compounds. Principessa farfallina Dr. Christa (alias Cristina) Mari for teaching me the most important Italian words and phrases and for organizing so many BBQs. Jeannine Hess for her contagiously happy attitude.

Philipp, Helena, Miriam, Richard, Silke, Anastasia, Sofia and Susi for the proof-reading of my thesis and feedback.

The members of the former ACI for a helpful and nice working atmosphere. Especially the ACI party crew for the fun and regular resets. Nathalie Melunsky-Fichter and especially Ramona Erni for the administrative support. Manfred Jöhri, Dr. Ferdinand Wild and Hanspeter Stalder for technical support. The MS-Service teams in the former ACI and OCI.

Anna, Magda, and Philipp for great moments and psychological support at the ISABC in Guangzhou, China (Ritz Carlton).

The Graduate School CMSZH for the funding, the aperos and the retreat. Especially Dr. Sabine Stockhause, Beatrice Spichtig for administrative support, and the best retreat orga team ever: Helena, Sandro Olveira, Dr. Georg Meseck, Dr. Mandes Schönherr and Dr. Daphne Diemer for a great time.

My friends not mentioned yet for their mental support, especially Andrea J., Kristin W., Franzi M., Nina M. and Dr. Achimi. And Klemens K. and Inga J. for a fantastic world championship. Schland ist Weltmeister!

My family for the amazing continuous support and for being always interested in what I'm doing: my parents Jürgen and Brigitte, my grandma Elfriede, my brother Frank, Bärbel, Lea and Fabio.

My love Philipp for everything. For taking me to Zurich. For all the support scientifically and psychologically. Correcting my drafts, listening to all my scientific talks and to all the small and bigger problems every day.

Table of Content

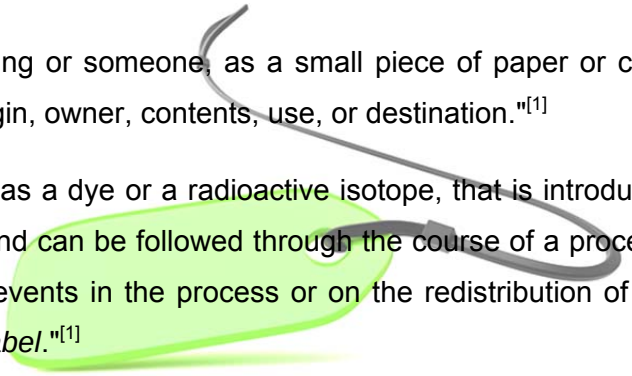
List of Abbreviations

ACN	Acetonitrile
AdoCbl	Adenosylcobalamin
AEEA	[2-(2-(Amino)ethoxy)ethoxy]acetic acid
ALEX	Alternating laser excitation
AqCbl	Aquocobalamin
ATP	Adenosine 5'-triphosphate
BB	Bromphenol blue
Bhoc	Benzhydryloxycarbonyl
BHQ	Black hole quencher
Boc	*tert*-Butyloxycarbonyl
bp	Basepairs
Cbl	Cobalamin
COSY	Correlation spectroscopy
CTP	Cytosine 5'-triphosphate
Cy	Cyanine
DAD	Diode array detector
DCC	*N*,*N'*-Dicyclocarbodiimide
DCM	Dichloromethane
DIPEA	*N*,*N*-Diisopropylethylamine
DMB	Dimethylbenzimidazole
DMF	Dimethylformamide
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DSC	*N*,*N'*-Disuccinimidyl carbonat
EBS	Exon binding site
EDTA	Ethylenediaminetetraacetic acid
en	Ethylenediamine
eq	Equivalents
ESI-MS	Mass spectrometry
EtBr	Ethidiumbromide
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Förster resonance energy transfer
GTP	Guanine 5'-triphosphate
HATU	1-[Bis(dimethylamino)methylene]-1*H*-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate

HDV	Hepatitis delta virus
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
HMBC	Heteronuclear multiple bond correlation
HMM	Hidden-Markov modeling
HPLC	High performance liquid chromatography
HR ESI-MS	High resolution electrospray ionization mass spectrometry
HSQC	Heteronuclear single quantum coherence
IBS	Intron binding site
K_D	Dissociation constant
LB	Lysogeny broth
LC	Liquid chromatography
LNA	Locked nucleic acid
Lys	Lysine
MALDI	Matrix-assisted laser desorption/ionization
MeCbl	Methylcobalamin
MOPS	3-(*N*-Morpholino)propanesulfonic acid
Mtt	Methyltrityl
NaOAc	Sodium acetate
NHS	*N*-Hydroxysuccinimide
NLS	Nuclear localization sequence
NMR	Nuclear magnetic resonance
nt/nts	Nucleotide, nucleotides
NTP	Nucleoside 5'-triphosphate
OD	Optic density
OHCbl	Hydroxocobalamin
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PLPG	Photo-labile protecting group
PNA	Peptide nucleic acid
$preQ_1$	Prequeuosine-1; 7-aminomethyl-7-deazaguanine
r.t.	Room temperature
RBS	Ribosomal binding site
RhB	Rhodamine B
RNA	Ribonucleic acid
RNAP	RNA polymerase

RP	Reverse phase
SAM	*S*-Adenosylmethionine
Sc.	*Saccharomyces cerevisiae*
SD	Shine-Dalgarno
sm	Single molecule
S_N2	Nucleophilic substitution
SPPS	Solid phase peptide synthesis
STO	Single turnover
TDP	Transition density plot
TEAA	Triethylammonium acetate
TFA	Trifluoroacetic acid
TIRF	Total internal reflection fluorescence
TIS	Triisopropylsilane
T_m	Melting temperature
TPP	Thiamine pyrophosphate
t_R	Retention time
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
Trolox	6-Hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid
U (MBU)	Molecular biology unit
UTP	Uridine 5’-triphosphate
UTR	Untranslated region
UV	Ultra violet
Vis	Visible
XC	Xylene cyanol

1 Introduction



label: "An item used to identify something or someone, as a small piece of paper or cloth attached to an article to designate its origin, owner, contents, use, or destination."[1]

tracer: "An identifiable substance, such as a dye or a radioactive isotope, that is introduced into a biological or mechanical system and can be followed through the course of a process, providing information on the pattern of events in the process or on the redistribution of the parts or elements involved. Also called *label*."[1]

Fluorescence labeling of molecules is a powerful method to elucidate processes hidden in the dark. It is very sensitive and allows investigating biological pathways in their natural environment. Either a fluorescent molecule can be tracked on the whole, or particular positions within a macromolecule by following the fluorescence emission. This gives information about positions, distances and dynamics. However, many challenges arise in the labeling of complex macromolecules. Several requirements have to be fulfilled by the label, such as specificity of the position, good yields, a fast and simple attachment procedure, no disturbance of the target and its function, stability, and visibility (brightness, signal-to-background). These aspects typically need to be optimized for each individual system.

This work addresses the topic of fluorescence labeling and deals with the development of new strategies to elucidate particular mechanisms of complex RNAs.
In the following chapter the basic characteristics of RNA are briefly introduced as well as FRET, a fluorescence technique used in these studies.

1.1 RNA

1.1.1 The Central Dogma of Molecular Biology

Deoxyribonucleic acids (DNA), ribonucleic acids (RNA) and proteins play a major role in every living organism. In the last 100 years the research in this field made outstanding efforts to clarify the structure, mechanism and interplay of these essential biological macromolecules. More than 30 Nobel prizes were awarded in this field up to now. But many questions are still unanswered and many details are not yet discovered. The interaction of DNA, RNA and proteins was summarized in 1958 to the “central dogma of molecular biology” (Figure 1).[2]

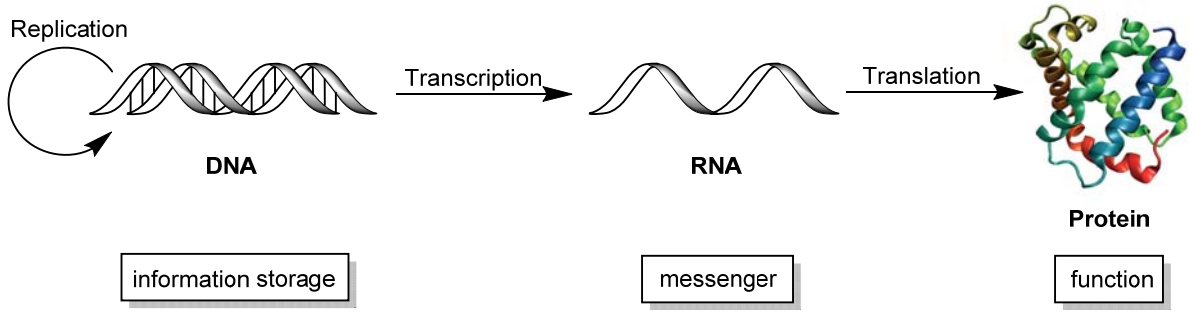


Figure 1: Central dogma of molecular biology. Transfer of genetic information from DNA to RNA to proteins.

DNA is the storage unit for genetic information in every living cell. It can be replicated and serves as a template for the synthesis of RNA. This process is called transcription. RNA can then further be translated into proteins, which are the essential machinery in living organism. However, it has been discovered that these general processes are much more complex and this scheme represents merely the most basic interactions of DNA, RNA and proteins.

1.1.2 Structure of RNA

The nucleic acids DNA and RNA contain the nucleobases guanine (G), adenine (A), cytosine (C) and in DNA thymine (T), or in RNA uracil (U). The bases have either purine based structures in case of A and G or pyrimidine based structures (Figure 2 a). The next bigger units (nucleotides) are formed together with a monosaccharide sugar, deoxyribose for DNA and ribose for RNA, and a phosphate group. These nucleotides are covalently linked between the sugar of one nucleotide and the phosphate group of the next nucleotide forming a highly negatively charged sugar-phosphate backbone. The asymmetric ends of these strands are denominated 3’ and 5’ according to the free linking position of the last nucleotide. The 3’ end has a terminal hydroxyl group on the sugar moiety and the 5’ end has a terminal

phosphate group (Figure 2 b). Most importantly, the nucleobases can additionally form intermolecular hydrogen-bond interactions following Watson-Crick base pairing rules (Figure 2 c).

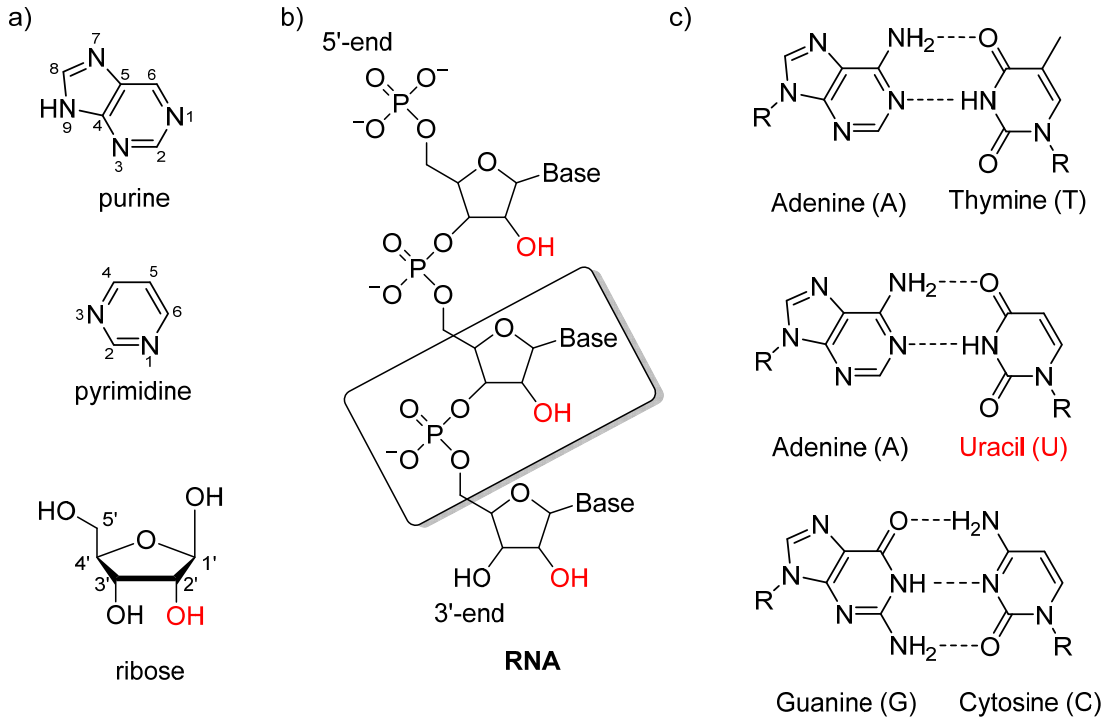


Figure 2: Characteristic structural elements of nucleic acids. Differences of RNA to DNA are marked in red. a) Basic structural elements with atom numbering. b) Covalently linked nucleotides (boxed) form an RNA single strand. c) Watson-Crick base pairing formed by the nucleobases.

Apart from the Watson-Crick base pairing scheme nucleic acids can pair in a Hoogsteen fashion (Figure 3) with the N7 of the purine base and the C6 group binding to N3 and N4 face of the pyrimidine base.[3,4] The base pairing interactions lead to a double helical structure of DNA formed by two complementary single strands.

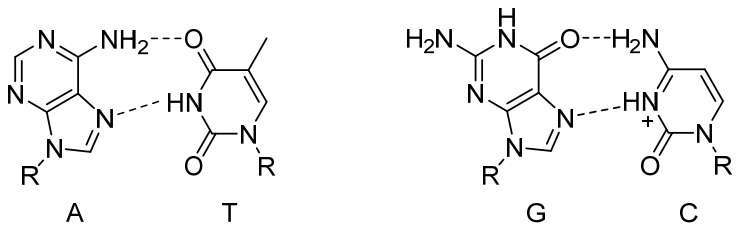


Figure 3: Hoogsteen base pairing. Purine N7 and C6 positions interact with pyrimidine N3 and N4 positions.

RNA typically appears single-stranded and can thus adopt various intrastrand structures. Besides Watson-Crick base pairs it can form further base pairs such as a G●U wobble pair.[5] This variety of interactions allows RNA to fold flexibly into a vast range of specific three-dimensional structures. RNA differs from DNA in the presence of a 2'-hydroxyl group at the ribose sugar (Figure 2 a and b). This not only makes the RNA less stable than DNA due to

hydrolysis, but also leads to important structural differences. RNA adopts a so-called A-form helix for double-stranded regions, whereas in DNA the B-form predominates.[6–8] Both helix types are right-handed but differ in the conformation of the sugar. This leads to a narrower major groove and shallower minor groove of A-from RNA compared to B-form DNA. This is important in terms of interaction with binding partners because the bases are differently accessible.

Besides the helical tertiary structures, single-stranded RNA can form a variety of characteristic secondary structure motifs such as hairpin loops, tetraloops, bulges, internal loops, multihelix and junction loops. These sub-structures can interact and play an important role for the tertiary structure of RNAs and hence their functions. An overview of several RNA motifs occurring in a functional RNA riboswitch aptamer (Chapter 2) is shown in Figure 4.

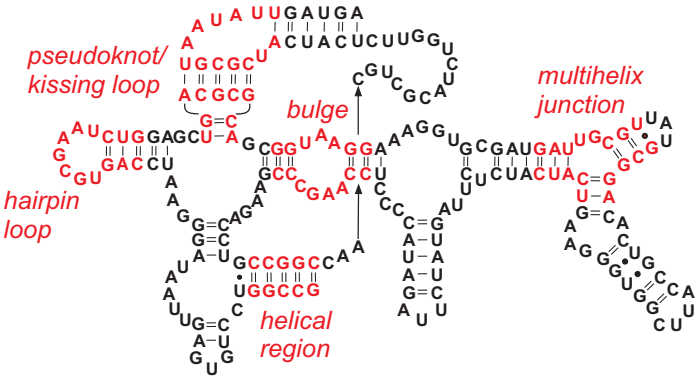


Figure 4: RNA secondary structure motifs exemplified by a riboswitch aptamer. Examples for different structure motifs are highlighted in red.

One example for such interplay is the ‘kissing interaction’ between two hairpin loops. The unpaired nucleotides of each loop base pair to form a so called loop-loop pseudoknot. Those structure motifs often play a crucial role in the functions of RNA as they can dynamically change in response to stimuli such as the binding of a metabolite.[9]

1.1.3 Folding and Functions of RNA

RNA is an essential biomolecule and extremely manifold in its occurrence and functions. In fact, the discovery that RNA fulfills such a variety of functions led to the 'RNA world hypothesis', proposing that RNA molecules were the precursors of all current life.[10,11] The RNA functions strongly depend on its structure, the so-called 'structure-function relationship'. The elucidation of RNA folding processes and mechanisms is therefore of interest for a better understanding of the associated functions.

To stabilize secondary and tertiary structures, RNA, and for that matter also DNA, requires cations such as mono- or divalent metal ions due to their negatively charged backbone. A pre-organization is reached in the presence of K^+ or Na^+ whereas Mg^{2+} is crucial for folding

into complex and functional tertiary structures. The metal ions induce a contraction or 'collapse' of the secondary structure into compact conformations.[12,13] The majority of metal ions interact non-specifically with the RNA, but some metal ions bind specifically to certain sites within an RNA structure and can also play crucial roles in RNA catalysis.[14]

The first function known for RNA was its role as messenger of information stored in DNA for the transfer into protein synthesis (Chapter 1.1.1), but since then a vast variety of other functions was discovered. RNA is involved in catalysis and regulation of gene expression, and can be a part of large protein-RNA complexes, for example ribosomes and spliceosomes. Different subtypes of RNA are categorized according to their function such as messenger (mRNA), transfer (tRNA), ribosomal (rRNA), small nuclear (snRNA), or small interfering (siRNA). Apart from mRNA all other kinds are classified as non-coding RNAs, *i.e.* functional but not translated into a protein.

The complexity of RNA functions makes it a fascinating molecule to study. To this end, non clarified folding mechanisms have to be understood. Not least, the elucidation of RNA folding processes and their three-dimensional architecture could lead to very practical applications in modern biotechnology and medicine, *e.g.* as novel targets in antibiotic drug development.[15]

1.2 FRET (Förster Resonance Energy Transfer)

The classical methods to elucidate the structure and function of RNA molecules, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy, have certain limitations. The crystallization of large and complex biomolecules is tedious and often unsuccessful. X-ray crystallography can provide detailed structural information, but cannot reveal dynamics of a process. Structures obtained from NMR measurements exhibit, besides the structure in solution, also information about structural dynamics. However, solving large RNA structures consisting of several hundreds of nucleotides is to date not possible. Single-molecule FRET for example can be used to determine distances within a molecule in motion to obtain conformational and kinetic details of folding processes. The fluorescence based techniques thus cannot be used for a complete structure determination but give the opportunity to investigate fundamental processes of large molecules and their dynamics in their natural environment. Fluorescence is manifold in its application and is extensively used in biotechnology, medical diagnostics, DNA sequencing, forensics and many more fields.

1.2.1 Principle

Luminescence is the emission of light from an electronically excited state of a molecule. The excitation is usually accomplished by irradiation with light. Depending on the nature of the excited state the emission process is called fluorescence or phosphorescence. Absorption and emission processes can be summarized in a Jablonski energy scheme (Figure 5).

Different electronic states such as singlet (S) and triplet (T) of a molecule are depicted with their vibrational energy levels (0, 1, 2). Upon light absorption the molecule is excited from the singlet ground state S_0 to an excited state S_1 or S_2 followed by a non-radiative rapid (10^{-12} s) relaxation to the lowest vibrational level of the first excited state with the same multiplicity, so-called internal conversion (ITC). Only from this state the molecule can return effectively to the ground state under emission of fluorescence, as stated by Kasha's rule. Alternatively, it can undergo a spin conversion to the first triplet state T_1 via intersystem crossing (ISC). Emission from T_1 is named phosphorescence and is shifted to longer wavelength compared to fluorescence. Transitions from T_1 to the singlet ground state S_0 are quantum mechanically forbidden and the emission rates appear rather slow (10^3 to 10^0 s^{-1}). In comparison, the emission rates of fluorescence are typically 10^8 s^{-1} corresponding to a fluorescence lifetime in the nanosecond range.

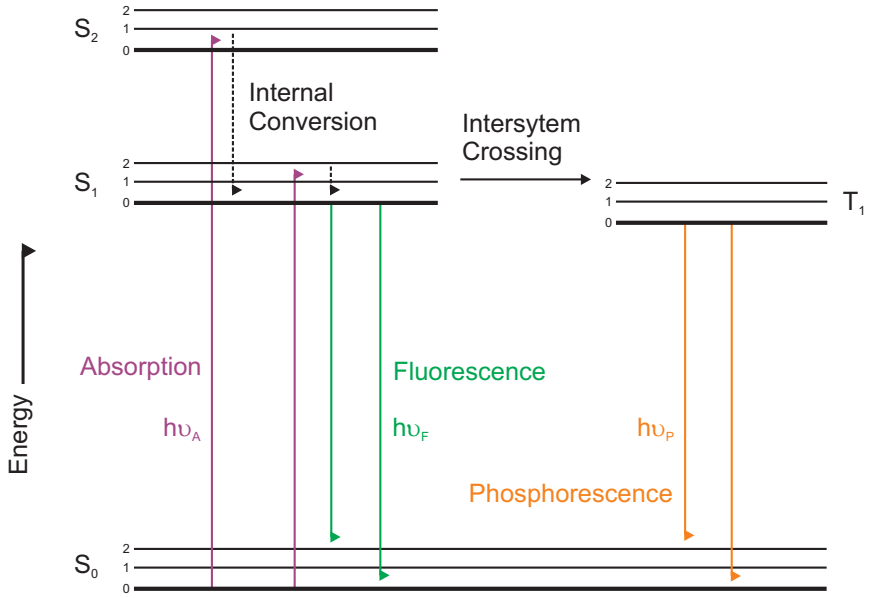


Figure 5: Jablonski energy scheme. Excitation from singlet ground state (S_0) to excited states (S_1, S_2), and emission of fluorescence or phosphorescence.

The number of emitted photons relative to the number of absorbed photons is expressed in the quantum yield, which is an important characteristic of a fluorophore. The higher the quantum yield is, the brighter is the emission. Various processes compete with the emissive decay and can therefore decrease the fluorescence quantum yield, *i.e.* quench the fluorescence. The excited fluorophore can for example be deactivated by interaction with another molecule in solution. Förster Resonance Energy Transfer (FRET) is a non-radiative energy transfer between two fluorescent dye molecules.[16] The energy is transferred from a donor, which is in its excited state, to an acceptor, which is in its ground state (Figure 6). For this process to occur, the emission spectrum of the donor has to overlap with the absorption

spectrum of the acceptor (compare Figure 9). The thus excited acceptor then relaxes via radiative emission.

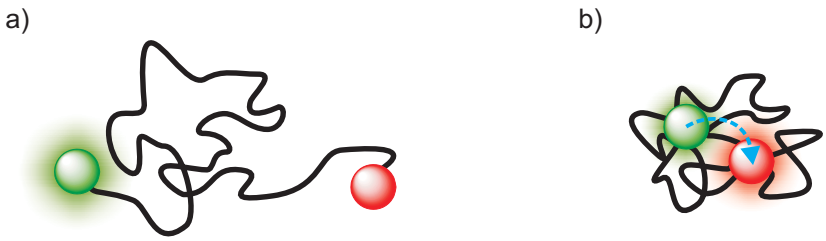


Figure 6: Energy transfer from an excited donor fluorophore to an acceptor. a) Excitation of a donor fluorophore (green) and florescence emission of the donor while bound to a macromolecule. b) Energy transfer from an excited donor fluorophore (green) to an acceptor (red) in close proximity. The emission intensity of the donor is decreased and the one of the acceptor is increased.

FRET is a highly distance dependent process. The efficiency of the energy transfer (E) can be expressed as a function of the distance between the dyes in accordance to equation (1):

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \tag{1}$$

with R = inter-dye distance and R_0 = Förster radius. The Förster radius R_0 is the distance at which the efficiency is 50 % (see Figure 7).

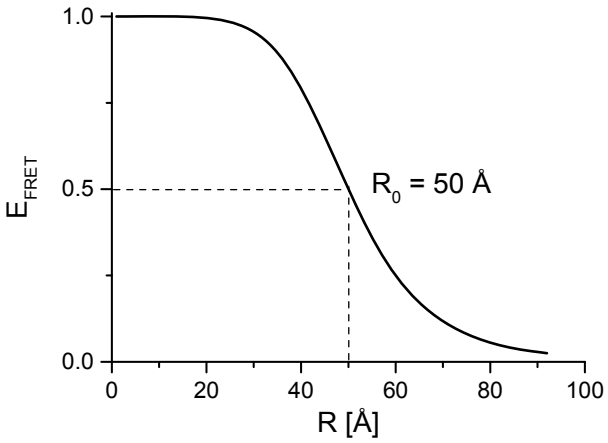


Figure 7: Distance dependency of the energy transfer efficiency E. R_0 is the Förster distance at which the efficiency is 50 %.

Typically Förster radii are in the range from 15 to 60 Å. Exploiting this distance dependency, a common application of FRET is the measurement of distances within macromolecules (spectroscopic ruler), between donor and acceptor dyes. If the dyes are attached to remote locations within a macromolecule, the monitoring of the FRET efficiency can give real-time information about conformational dynamics of the molecule.

Importantly, the efficiency of the energy transfer depends on several factors:

- relative emission strength of the donor and acceptor fluorophores,
- donor-acceptor separation (distance),
- fluorescence quantum yield of the isolated donor,
- relative orientation of donor and acceptor transition dipoles,
- spectral overlap between donor emission and acceptor excitation.

Only the donor-acceptor distance directly appears in equation (1). The donor quantum yield, relative orientation, and spectral overlap are indirectly included as they influence the Förster radius R_0. The Förster radius (in Å) can thus be expressed as:

$$R_0^6 = 8.79 \times 10^{-5}\left(\kappa^2 n^{-4} Q_D J(\lambda)\right) \quad (2)$$

with κ^2 being a factor describing the relative orientation of the fluorophore dipoles, Q_D being the quantum yield of the donor in absence of an acceptor, n being the refractive index, and $J(\lambda)$ being the overlap integral over the wavelength in nm, which expresses the degree of spectral overlap between donor emission and acceptor absorption.[17]

The Förster radius R_0 is a specific constant for a given dye pair. However, it is based on the assumption of free rotation of the fluorophores. The attachment of the dye to a macromolecule can impede the rotational freedom, leading to a changed apparent R_0 value.

As can be seen on the basis of the discussion about the Förster radius, the choice of the dyes is of high importance. Many additional factors have to be considered such as photostability, availability, and interaction with the target molecule.

1.2.2 Cyanine Dyes

Cyanine dyes (Cy) are widely used in biological applications such as structural investigations of nucleic acids. A common FRET pair is Cy3 (donor) and Cy5 (acceptor). The popularity of these cyanine dyes is due to their photostability and commercial availability.[18] The class of (carbo)cyanine dyes are based on a polymethine structure, which is a CH chain with alternating single and double bonds giving an extended π-conjugates system (see Figure 8). Generally, an extension of the π-system leads to a decrease in the HOMO-LUMO gap and thus to a red-shift of the absorption band. In this way, the emission and excitation wavelength can be tuned. The number in the name of the cyanine dyes designates the length of this chain. Cyanines typically have the structure depicted in Figure 8, but variations in the substituents R (*e.g.* methyl, ethyl) and R' (*e.g.* sulfonate, H), and even more drastically substitutions such as the replacement of the phenyl by naphthyl moieties are possible.

For the present studies Cy3 (n = 1) and Cy5 (n = 2) with R = methyl, R' = H or SO_3^{2-} were used.

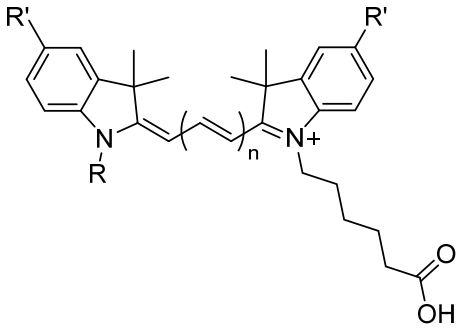


Figure 8: Common structure of cyanine dyes. n = 1 Cy3, n = 2 Cy5. The number in the name describes the length of the polymethine chain.

Cyanine dyes have absorption and emission wavelengths above 550 nm. They typically display a rather small Stokes shift with the emission maxima being red-shifted about 30 nm from the absorption maxima (Figure 9).[17]

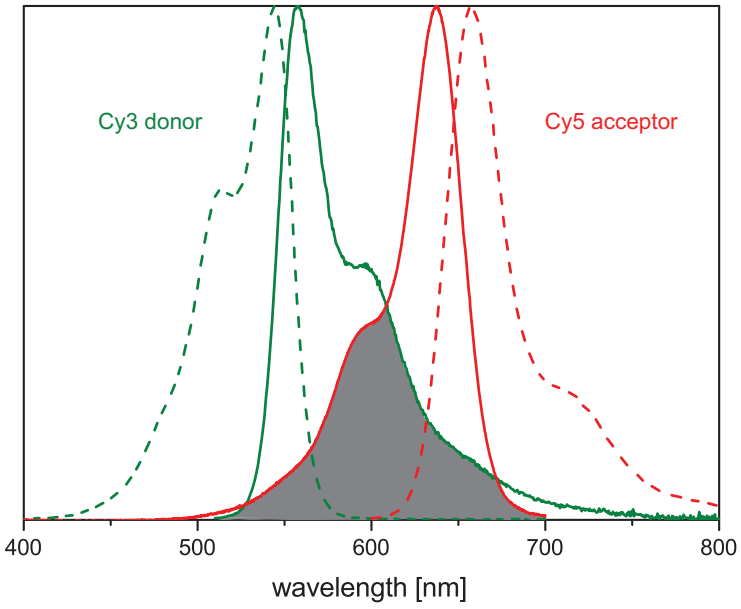


Figure 9: Absorption and emission spectra of Cy3 and Cy5. Overlay of normalized absorption and emission spectra of Cy3 (green) and Cy5 (red).[19] The spectral overlap of the Cy3 emission with the Cy5 absorption (solid lines) is depicted in grey.

The photophysical properties of the cyanine dyes depend strongly on their surroundings. For example, photoisomerization, that is rotation among the polymethine chain, deactivates the first singlet excited state to a ground state photoisomer. The efficiency of this process is dependent on the environment,[20–22] leading to a strong dependence of the their quantum yields on the viscosity of solvent.[23]

A variation on the classical fluorophore-fluorophore energy transfer scheme is the use of so-called dark quencher dyes, which can function as acceptors. They dissipate the absorbed energy without exhibiting fluorescence, therefore reducing issues due to background fluorescence. One example for dark quenchers are the so-called 'black hole quenchers' (BHQ), which quench by FRET and static quenching mechanisms.[24]

Cyanine Dyes as Fluorescent Probes for Nucleic Acids

Cyanine dyes are applied in various fluorescence techniques, *e.g.* for the imaging of single molecules in living cells,[25] and fluorescence *in situ* hybridization (FISH).[26] Especially Cy3 and Cy5 are often used for covalent labeling of nucleic acids, proteins and peptides.[27]
For the labeling of nucleic acids a couple of important points have to be considered. The attachment of the dye to the RNA, and macromolecules in general, can impede the rotational freedom. Increasing the length of linkage between the dye and the RNA can support the free dye rotation but leads to greater uncertainties for intermolecular distances in RNA due to the increased conformational flexibility of the linker.[21] Moreover, it is known that the fluorescence of Cy3 and Cy5 is enhanced by the attachment to nucleic acids. This is due to a reduced *trans-cis* isomerization (*vide supra*) caused by dye-nucleobase interactions.[20–22,28,29] Furthermore, terminally attached Cy3 and Cy5 appear stacked at the 5'-end of a DNA duplex similar to an additional base pair. This behavior has been observed with both a short (3 carbon) and a long (13 atoms) tether. The stacking thus influences the FRET efficiency.[30–34]

1.2.3 Single-Molecule FRET Microscopy

Förster Resonance Energy Transfer is a powerful tool to study the conformational dynamics of biomolecules, especially on the single molecule level (smFRET).[35–38] Following a single molecule at a time with FRET reveals detailed information about folding pathways that are lost in ensemble measurements due to the inherent signal averaging. Several fundamental processes have been investigated with this technique, *e.g.* the dynamics of ribosomal translation,[39] and the ligand recognition of a riboswitch aptamer.[40]
More generally, single-molecule microscopy is used in a variety of different approaches such as fluorescence correlation spectroscopy, atomic force microscopy, confocal microscopy or total internal reflection fluorescence (TIRF) microscopy. TIRF is based on the excitation of molecules by an evanescent wave, which is formed at the interface between the microscope slide and the sample solution and only extends 100 – 200 nm in the solution. The small excitation volumes lead to a reduced background if surface-immobilized molecules are imaged. The TIRF effect can be realized by a prism-based (Figure 10) or objective-based method. A conventional TIRF setup consists of laser(s) for the excitation of the fluorophores, a measuring chamber placed on a microscope with an objective to collect the fluorescence

emission, and a CCD camera for the detection of the single-molecule fluorescence. Notably, in all setups featuring a camera the time resolution is limited by the frame rate of the CCD camera.[18,41]

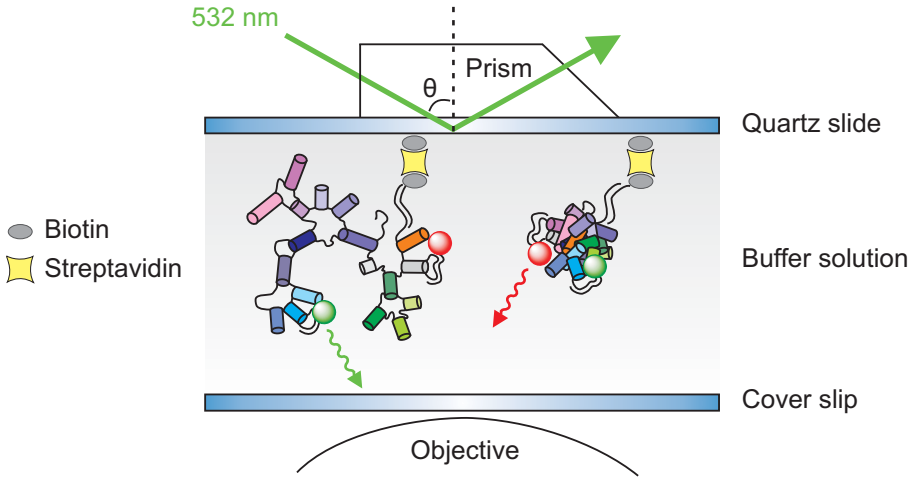


Figure 10: Single-molecule detection by prism-based TIRF microscopy. The laser beam (green line) is focused through a prism towards immobilized molecules labeled with dyes. The angle of incidence θ must be greater than θ_c (critical angle) to achieve TIRF. Emitted fluorescence (red and green wavy arrows) is detected by an objective. *Figure based on Ref. [42]*.

Each well separated single molecule can be detected as individual spot in wavelength separated donor and acceptor channels to obtain fluorescence trajectories. In case of energy transfer changes, these trajectories typically exhibit an anti-correlated behavior. Due to the distance dependency of the energy transfer between the fluorophores, the intensity of donor emission is decreased while the one of the acceptor is increased when the fluorophores get closer to one another (Figure 11 a). Apparent FRET is thus calculated (Figure 11 b) from the intensities of the donor (I_D) and acceptor (I_A) emission:

$$FRET = \frac{I_A}{I_D + I_A} \tag{3}$$

The obtained FRET is further analyzed by specialized software which allows a detailed thermodynamic and kinetic analysis. FRET trajectories can be binned into histograms to reveal the relative probability of each state (Figure 11 c) and rate constants can be determined from the dwell-times of different states (Figure 11 b).

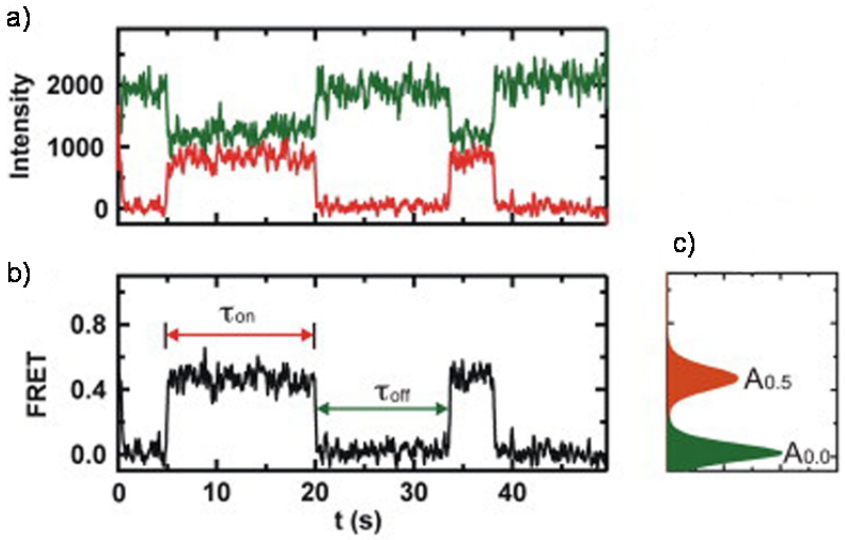


Figure 11: Example of single-molecule trajectories and data analysis by use of an RNA model system. a) Fluorescence emission time trajectories from donor (green) and acceptor (red) fluorophores. b) Calculated FRET time trajectory with dwell-times (τ) of formation and dissociation states of an RNA complex. c) FRET histogram obtained from binning of single time trajectories. *Reprinted from Ref. [41], Copyright (2009), with permission from Elsevier.*

Sample immobilization can be reached by different methods. A commonly used approach is the use of a biotin-streptavidin linkage (Figure 10). In case of nucleic acid studies biotinylated bovine serum albumin (BSA) is adsorbed at the surface of a quartz slide. For proteins polyethylene glycol (PEG) is often used for passivation.[18] The photostability of the fluorophores can be increased by addition of different reagents such as oxygen scavenging systems. Additionally, triplet-state quenchers can be added, *e.g.* Trolox, to reduce photoinduced blinking, *i.e.* random fluorescence intensity-fluctuations, of the dyes.[18,35,41,43–45] As smFRET is an extremely sensitive method reliable results depend on a vast variety of factors such as sample preparation, selection of the fluorophores and their respective position within the target molecule, background noise issues, photostability, and data analysis. However, the opportunity to answer fundamental questions about transcription, translation, RNA folding and catalysis, conformational changes and many more makes this method so important.

1.3 Thesis Outline

The aim of this work is the development and improvement of new fluorescence labeling strategies for the investigation of folding mechanism of large functional RNA systems by Förster Resonance Energy Transfer. Our particular interest is the elucidation of the complex mechanism of B_{12} binding to the *btuB* riboswitch of *E. coli* and its subsequent structural rearrangement (Figure 12). As FRET is based on the energy transfer from a donor fluorophore to an acceptor, a labeling of the target molecule(s) is necessary.

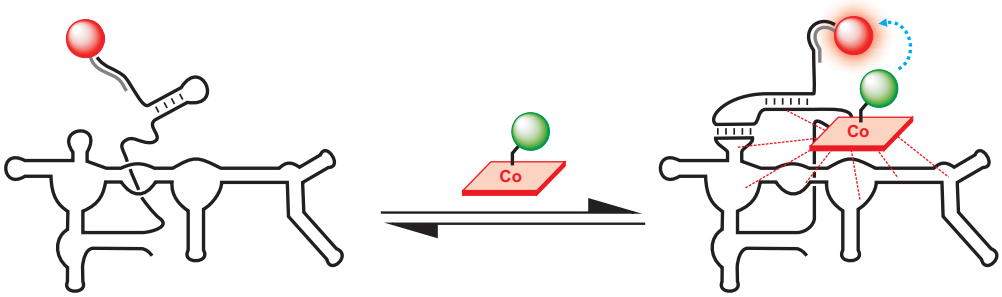


Figure 12: Labeling strategy for the investigation of the *btuB* riboswitch by FRET. Structural rearrangement of the RNA upon binding of B_{12} and resulting FRET between the fluorophores (green and red).

Chapter 2 focuses on the labeling of the metabolite B_{12}. The synthesis of new fluorescent B_{12} derivatives and their characterization is described. Furthermore, in-line probing experiments are conducted to characterize the influence of the modified derivatives on the *btuB* riboswitch structure.

Chapter 3 deals with the labeling of large RNAs with PNA-dye conjugates as specific indirect labels. The investigation and optimization of the labeling is described using the group II intron ribozyme as model. This system has been well studied and therefore allows a critical comparison of data obtained with the new PNA-based strategy and the previously used DNA-based one. The goal is to proof and improve the applicability of PNA probes for fluorescence techniques especially smFRET.

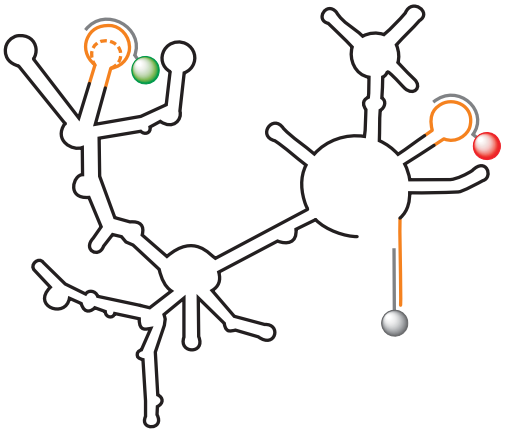


Figure 13: Group II intron model system with dye- labels.

Chapter 4 goes beyond the scope of the previous chapters and describes collaborations for further applications of the fluorescent PNA derivatives, as well as studies on the influence of Rhenium compounds on RNA, exemplified once more on the group II intron model.

2 Labeling of B_{12}

2.1 Introduction

The *btuB* riboswitch RNA of *Escherichia coli* possesses a gene regulating function by rearranging its three dimensional structure ("switching") upon specific binding of cobalamin (Chapter 2.3).[46] This B_{12}-dependent riboswitch has one of the most complex riboswitch structures and its metabolite B_{12} is the largest metabolite binding to RNA. Yet, the binding of the B_{12} to the *btuB* riboswitch and the structural rearrangement of the RNA is not fully understood. The elucidation of this complex mechanism can give new insights into fundamental processes of RNA function and interaction. Manifold applications could arise from this. Riboswitches are already under investigation for intracellular sensors for the detection of small molecule concentrations to study cellular processes.[47] Synthetic riboswitches can be used to control gene expression in cell-like systems.[48] Furthermore, riboswitches have been identified as highly selective antibacterial drug targets, addressable by non-functional metabolite mimics.[15]

The secondary structure and binding/switching capabilities of the RNA can be investigated by in-line probing (Chapter 2.3.5).[49] Moreover, several smaller riboswitches have already been explored by single-molecule FRET.[50] This technique could be used to gather information about the folding dynamics of the *btuB* riboswitch of *E. coli* upon metabolite binding. To date, this large and complex system has not been studied by such a real-time method.

2.2 B_{12}

2.2.1 Structure and Reactivity

Cobalamines, also named B_{12}, are complex biomolecules and among the largest metabolites, and play important roles in living organism, including the human metabolism. The most prominent B_{12} derivatives are vitamin B_{12} (cyanocobalamin, CNCbl, VitB_{12}), the organometallic cofactor coenzyme B_{12} (5'-deoxy-5'adenosylcobalamin, AdoCbl), methylcobalamin (MeCbl), aquocobalamin (AqCbl) and hydroxocobalamin (OHCbl). The red cobalt(III)-complex vitamin B_{12} was first isolated in the late 1940's. For the elucidation of the complex structure of B_{12}, and others, by X-ray crystallography D. Crowfoot Hodgkin was awarded the Nobel Prize in 1964 (Figure 15).[51]

The basic structure of B_{12} derivatives is composed of several characteristic subunits: a cobalt-coordinating corrin macrocycle with a dimethyl-benzimidazole (DMB) nucleotide moiety at the α-face (lower) and a ligand (R) on the β-face (Figure 14). The Co^{3+} center is coordinated to six ligand atoms in a pseudo-octahedral conformation including four nitrogen atoms in the corrin ring as part of the π-conjugated chromophore system.[52] These corrin based cyclic systems containing four pyrrole rings (A-D in Figure 14) similar to porphyrins belong to the group of corrinoids. So-called 'complete' corrinoids coordinate a nucleotide base at the α-face of the cobalt-center ('base-on'). Typically this 'base-on' form is favored and the one being recognized by proteins. However, a switch to a 'base-off' (DMB-base de-coordinated) form can occur at higher temperatures, upon protonation of the nucleotide base or substitution by an external group such as histidine by binding to proteins or cyanide. The molecular switch between 'base-on' and 'base-off' is an important feature of B_{12} to modulate its organometallic reactivity and redox chemistry regarding the needs of the corresponding enzyme.[53–55]

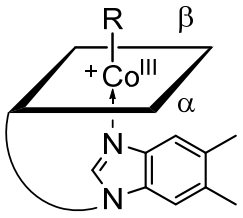


Figure 14: Schematic structure of cobalamins.

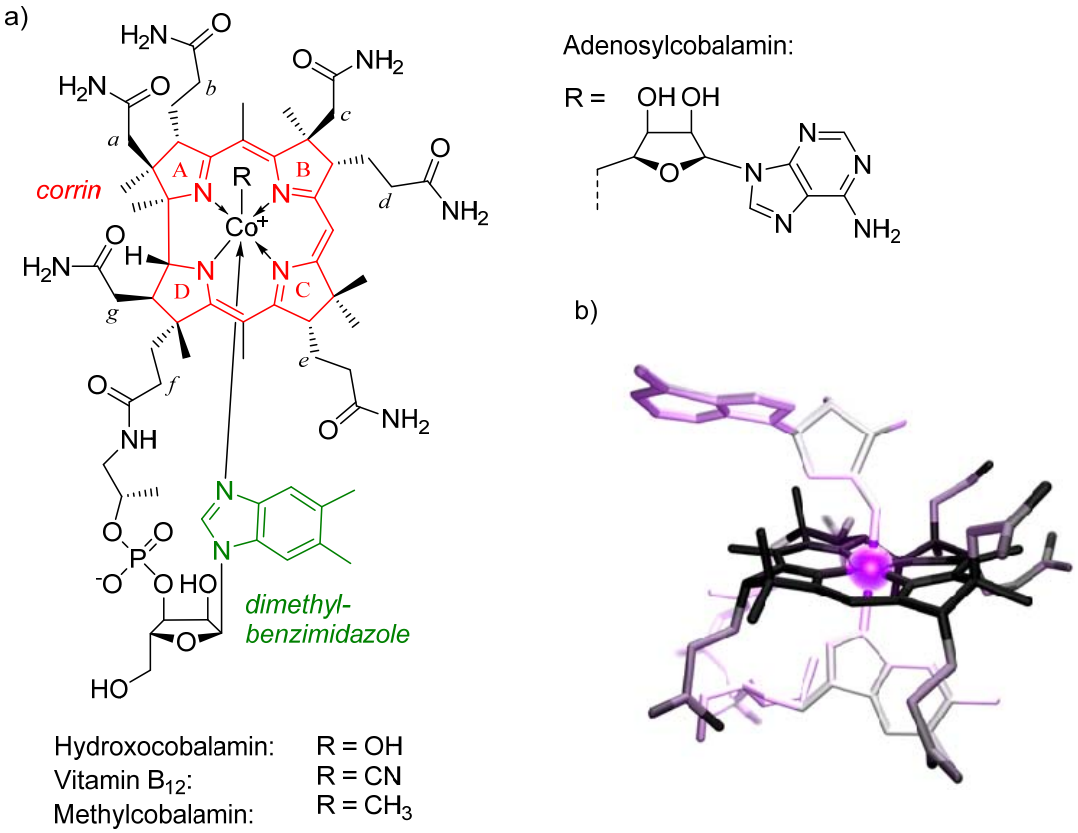
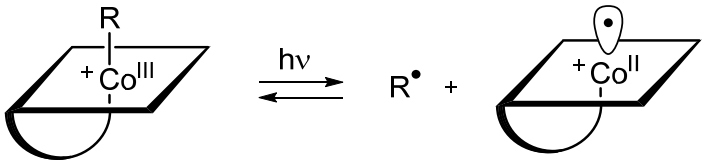


Figure 15: Structure and nomenclature of cobalamin derivatives. a) Chemical structure of cobalamin and characteristic elements.[56] b) Crystal structure of AdoCbl; *figure kindly provided by Dr. Joachim Schnabl*.

In coenzyme B_{12} (adenosylcobalamin, AdoCbl) the ligand R = 5'-deoxy-5'-adenosyl is coordinated at the β-face of the cobalt center via a highly light sensitive organometallic Co-C bond (Figure 15).[57,58] The Co-C bond can be cleaved homolytically (Scheme 1) via a one-electron 'inner-sphere' reduction as for AdoCbl or heterolytically as a two-electron 'inner-sphere' reduction as for methylcobalamin.[53,59,60] Homolytic cleavage of AdoCbl with a bond dissociation energy of ~30 kcal mol^{-1} leads to a formation of a 5'-deoxyadenosyl radical and cob(II)alamin (B_{12r}; Scheme 1).[61] The penta-coordinated radicaloid B_{12r} can further react with alkyl radicals as a β-specific 'radical trap' or with oxygen to form hydroxocob(III)alamin.[62–64]



Scheme 1: Photolytic Co-C bond cleavage. Formation of a radical ($R^{\bullet}$) and cob(II)alamin upon homolytic Co-C bond cleavage of organometallic B_{12} derivatives; *e.g.* R = 5'-deoxyadenosyl.[56]

MeCbl can react via a nucleophilic attack to form cob(I)alamin and a methylating agent. With its ability to transfer the methyl group to nucleophiles, MeCbl fulfills a cofactor function for enzymes.[53,56]

2.2.2 Biological Relevance

The physiological role of vitamin B_{12} derivatives is mainly fulfilled by the organometallic cofactors, which are involved in human, animal and microbial metabolism. In fact, B_{12} cofactors catalyze a variety of biological processes that are essential for many living organisms. Several enzymes, *e.g.* corrinoid dehalogenases and methyl transferases, use B_{12} as a cofactor, often in a 'base-off/His-on' form with a histidine provided by the 'B_{12}-binding motif' of the protein (His-X-Asp). B_{12} binds further to proteins such as B_{12} transport proteins and interacts with nucleic acids (Chapter 2.3.4).[53,65,66]

Microorganisms in the class of bacteria and archaea are able to partially or fully synthesize B_{12}. *E. coli* and several other bacteria require either cobinamides ('incomplete' corrinoids that lack the nucleotide) or different cobalamins for the conversion into MeCbl and coenzyme B_{12}. In gram-negative bacteria the corrinoid transport depends on the 'B-twelve-uptake' (Btu) system (Figure 16). The outer membrane transporter BtuB binds the corrinoids and transfers the substrate into the periplasm. It is trapped there by BtuF and further transferred to the cytoplasm via the inner membrane transporter $BtuC_2D_2F$.[56,67–72]

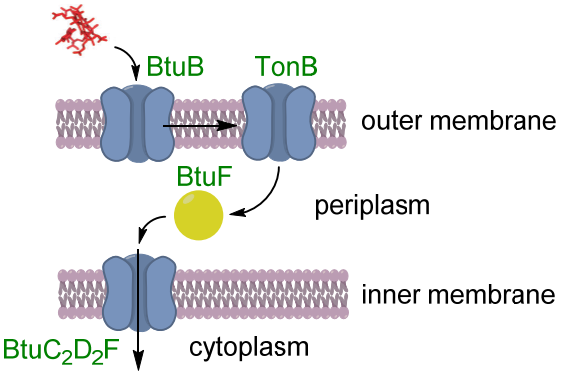


Figure 16: Schematic overview of B_{12} transport in *E. coli* by the Btu system. Corrinoids cross the outer membrane trough the transport protein BtuB, are transferred through the periplasm by BtuF and further to the cytoplasm by $BtuC_2D_2F$.

Mammals cannot synthesize corrinoids at all and are dependent on uptake by nutrition. The corrinoids such as vitamin B_{12} are then metabolized to the physiologically relevant cofactors, *e.g.* AdoCbl.

2.3 Riboswitches

2.3.1 Classification and Function

The relationship between structure and function of biomolecules is fundamental for the explanation of biochemical processes. An example for this is the class of riboswitches, regulatory elements that serve as *cis*-acting genetic control modules with response to target ligands.[73–76] Riboswitches are involved in the synthesis or import of cognate metabolites by feedback repression of genes, and recognize fundamental metabolites such as coenzymes, amino acids, nucleobases and aminosugars. More than 20 distinct classes of riboswitches have been identified to date. The size of the metabolite-binding domains can vary from smaller structures as for $preQ_1$ (34 nts, prequeuosine-1) to large complex structures as for AdoCbl (~200 nts). These metabolite-sensing RNAs are involved in regulatory mechanisms of transcription, translation and alternative splicing. Typically, riboswitches are located in the 5' untranslated region (UTR) of bacterial mRNA. They generally consist of an aptamer domain and an expression platform, and can up- or down-regulate gene expression. Depending on the metabolite concentration both domains participate in folding and in a structural rearrangement forming an intrinsic transcription terminator or competing RNA structure (Figure 17). The aptamer is crucial for the binding of the metabolite and is interacting with the expression platform that translates the information to control the expression.[73] Two main types of expression platforms are known. One is based on the mechanism of transcription termination. An intrinsic terminator stem is formed, typically followed by five to nine uridines, which causes the RNA polymerase (RNAP) to halt

transcription (Figure 17 a).[77,78] The other regulatory mechanism relies on the influenced accessibility of the Shine-Dalgarno (SD) sequence, which is a ribosomal binding site consisting of a short, purine-rich sequence that recruits 30S ribosomal subunits near AUG start codons of mRNA.[79] Riboswitches involving a SD sequence modulate translation initiation upon their ligand-induced structural changes (Figure 17 b).[73,74,80]

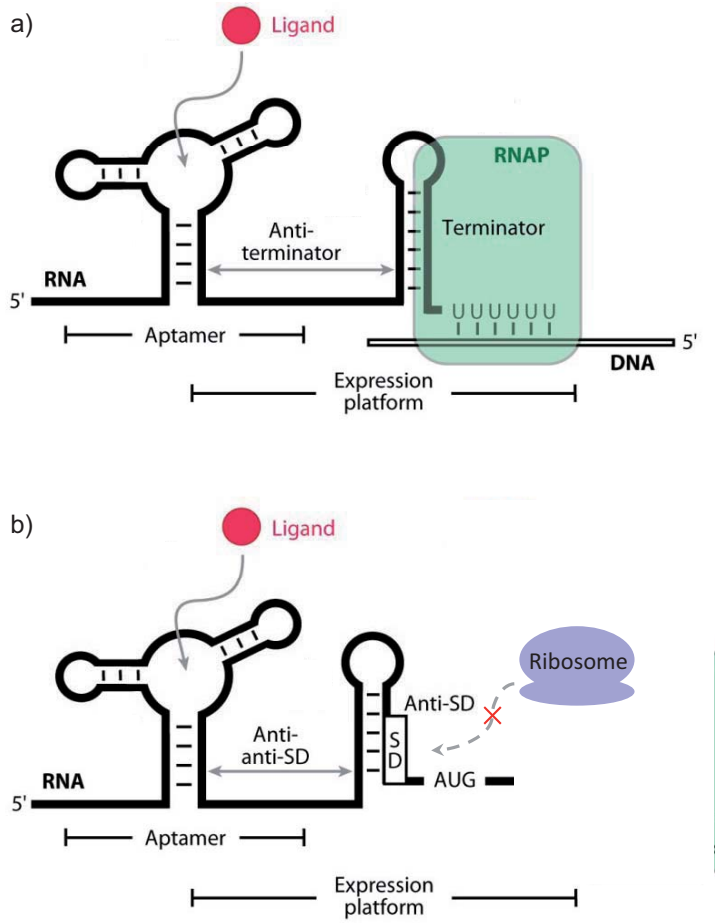


Figure 17: Mechanism of riboswitch-mediated gene control. Schematic representation of a riboswitch that represses gene expression upon ligand binding by controlling a) transcription termination, and b) translation initiation. *Figure adapted from Ref. [73]*.

2.3.2 B_{12}-Sensing Riboswitches

B_{12}-riboswitches are among the largest and most complex of these regulatory RNAs. They can bind B_{12} derivatives such as coenzyme B_{12}, methylcobalamin and aquocobalamin. Comparison among known riboswitch structures revealed common structural elements. Based on this, B_{12}-dependent riboswitches are divided into two classes depending on the B_{12} derivative they bind.[81–83] The AdoCbl-binding class differs from the AqCbl- and MeCbl-binding class in an additional peripheral extension of the pairing element P6 by P7-P11 (Figure 18 a). A common core around a four-way junction (J), and a 'kissing-loop' interaction (see Chapter 1.1) between loop L5 and L13 are similar in both classes. In transcription

regulating riboswitches, P13 typically involves the anti-terminator, whereas P13 in translation regulating riboswitches contains the ribosomal binding site (RBS) or related structures (*vide supra*). The junction J6/3 between the pairing elements P3 and P6 is the key element for cobalamin recognition, as known from crystal structures of AdoCbl-binding riboswitches of *Symbiobacterium thermophilum* and *Thermoanaerobacter tengcongensis* (*Tte*; Figure 18 b and c), as well as of the smaller AqCbl-binding *env8*AqCbl. The conformation of the junction J6/3 in the AdoCbl-binding class is defined by the extensions P7-P11 providing a cavity for the adenosyl moiety. It serves as a cobalamin-binding pocket mainly driven by van der Waals interactions. The AqCbl-binding-class reaches ligand specificity for AqCbl by the arrangement of the purines within the junction, limiting the space available for the B_{12}. Consequently, the binding of B_{12} derivatives containing large substituents on the β-face is prevented by sterical hindrance. The kissing-loop interaction L5-L13 encapsulates the bound cobalamin.[81–83]

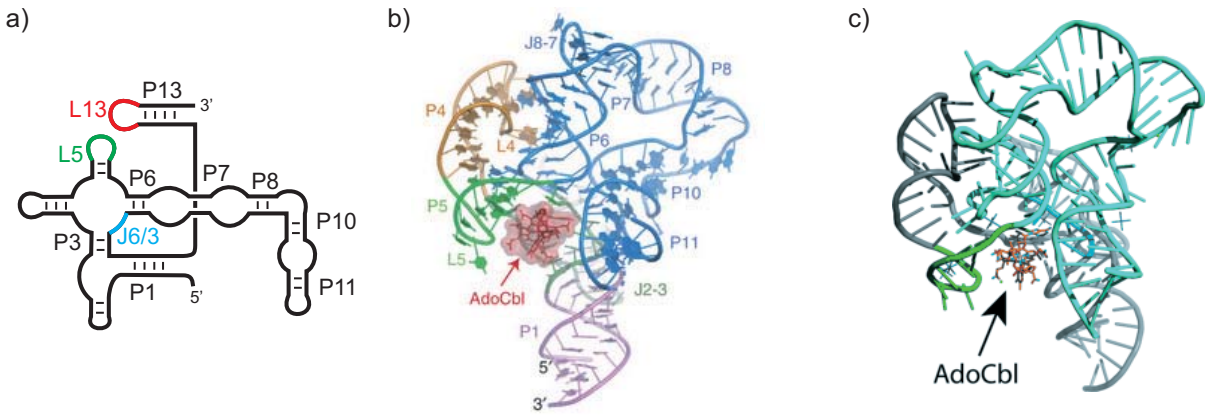


Figure 18: Crystal structures of B_{12} riboswitches. a) Secondary structure organization of AdoCbl riboswitches. Junction J6/3 (blue) is the key element for ligand recognition. Loops L5 (green) and L13 (red) form a kissing-loop interaction upon B_{12} binding.[81] b) Ribbon representation of the AdoCbl riboswitch structure of *S. thermophilum* bound to AdoCbl. *Figure adapted by permission from Macmillan Publishers Ltd: Nature Structural & Molecular Biology, Ref. [82], copyright 2012*. c) Representation of the AdoCbl-bound riboswitch *Tte*. *Figure adapted from Ref. [81] with permission*.

2.3.3 The *btuB* Riboswitch of *E. coli*

One of the first representatives in the class of riboswitches that was discovered is the *btuB* riboswitch of *Escherichia coli*. It specifically binds coenzyme B_{12} as its natural ligand. The *btuB* riboswitch is located in the 5' UTR of the *btuB* gene, which encodes an outer membrane protein that is involved in the B_{12} transport (Chapter 2.2.2). Expression of the *btuB* protein is repressed only by AdoCbl, by inhibiting ribosome binding to the riboswitch.[74,84] This riboswitch is therefore actively involved in the regulation of the cobalamin concentration.

Phylogeny, structure modeling, structural mapping and transcription pausing studies provided a consensus structure model of the *btuB* riboswitch of *E. coli*.[46,84,85] Nevertheless, the complete structure remains unknown as no crystal structure is available so far for this riboswitch. The riboswitch is composed of two functional domains; one serves as natural aptamer (~200 nts) that binds B_{12}, and the other one is an expression platform (~40 nts) that is responsible for the gene regulation (Figure 19 a).[46]

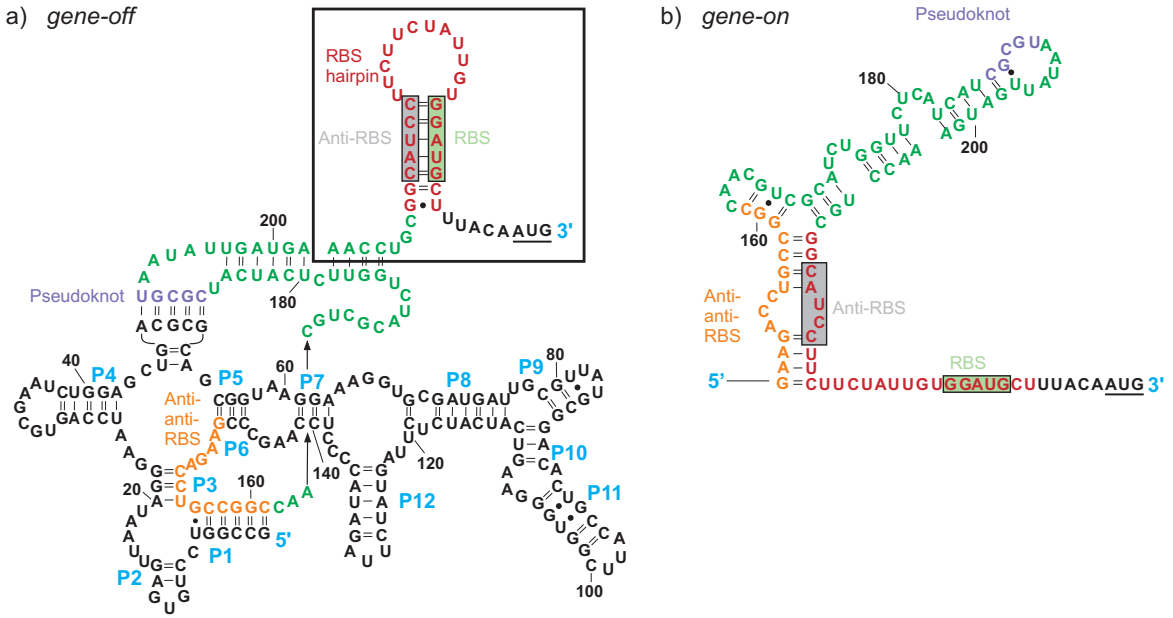


Figure 19: Secondary structure of the *btuB* riboswitch of *E. coli*. a) Aptamer and expression platform (box) in the gene-off state (B_{12} bound and ribosomal binding site (RBS) blocked). b) Antiaptamer, gene-on (B_{12} not bound, RBS accessible).The start codon AUG of the coding region is underlined.[46,85]

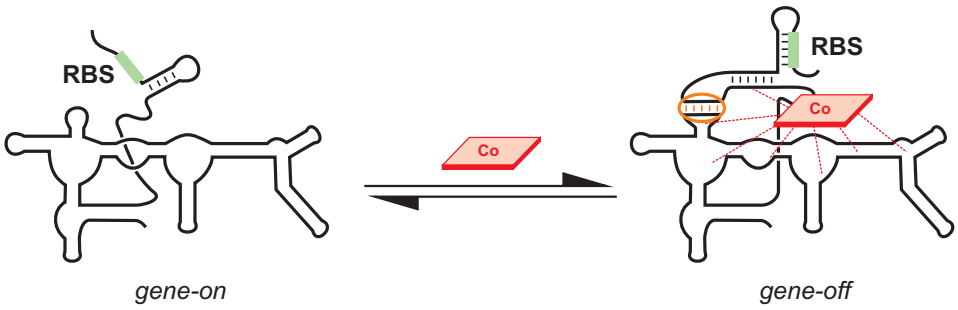


Figure 20: Folding of the *btuB* riboswitch of *E. coli* upon B_{12} binding. The ribosomal binding site (RBS) marked in green is blocked due to the structural rearrangement of the riboswitch upon B_{12} binding (red). A kissing-loop interaction is highlighted in orange. *Based on figure from Dr. Sofia Gallo*.[86]

At high cellular concentrations of coenzyme B_{12}, the metabolite binds to the aptamer and induces a conformational switch (schematically shown in Figure 20). The expression platform, including the ribosomal binding site (RBS), is thereby structurally rearranged. Consequently, the RBS becomes inaccessible and the translation of the protein is suppressed (bound, gene-off, Figure 19 a). At low concentration of the B_{12} metabolite, the

RNA adopts an antiaptamer structure that allows for translation (apo, gene-on, Figure 19 b). More in detail, the large aptamer domain folds into a complex tertiary structure depending on B_{12} binding, whereas the expression platform folds into simple secondary structures dictated by the aptamer. When B_{12} is bound the anti-anti-RBS region is part of the aptamer, the formation of a hairpin loop of the anti-RBS with the RBS region takes place and gene translation is thus turned off. In the ligand-free apo form the anti-anti-RBS is paired with the anti-RBS including a three-way junction, and the RBS is in a single-stranded region and consequently available for ribosome binding.[46,74,85] Experiments carried out in transcription mixture and in comparison with purified and refolded RNA revealed that the transcription plays a crucial role for efficient folding and conformational change of the *btuB* riboswitch. Transcriptional pausing by *E. coli* RNA polymerase at the end of the aptamer domain coordinates the formation of the secondary structure of the expression platform.[85]

2.3.4 Interactions of the *btuB* Riboswitch and Coenzyme B_{12} Derivatives

The conformational changes of the *btuB* riboswitch of *E. coli* upon binding of the metabolite AdoCbl can be monitored by in-line probing titration experiments (Chapter 2.3.5). With this method appearance and disappearance of distinct cleavage bands as a function of concentration are detected. From the literature, nine cleavage sites are known for the aptamer sequence (202 nts), which are affected by AdoCbl binding.[46,87] Other B_{12} derivatives are also recognized by the riboswitch, such as vitamin B_{12}, adenosyl factor A (AdoFactA) and adenosyl-cobinamide (AdoCobi). These derivatives lack either the upper organometallic group or the lower base. They induce the same conformational changes of the RNA but exhibit an up to 1000-fold lower binding affinity than AdoCbl. Consequently, apical ligands that are bound to the cobalt center are generally expected to only influence the binding affinity, but not the folding.[87] Further studies with modified side chains of the corrin ring revealed that the unchanged corrin moiety of cobalamin is crucial for the correct folding of the RNA.[88] Regardless of these general trends, a prediction of influences on the riboswitch folding behavior and binding affinity from other B_{12} derivative is difficult and in-line probing experiments are needed to verify hypotheses.

2.3.5 Principle of In-line Probing

The secondary structure of complex RNA molecules can be elucidated with the well-established in-line probing technique. Furthermore, ligand binding affinities can be determined. In particular, this assay can be used for the investigation of a riboswitch aptamer.[49]

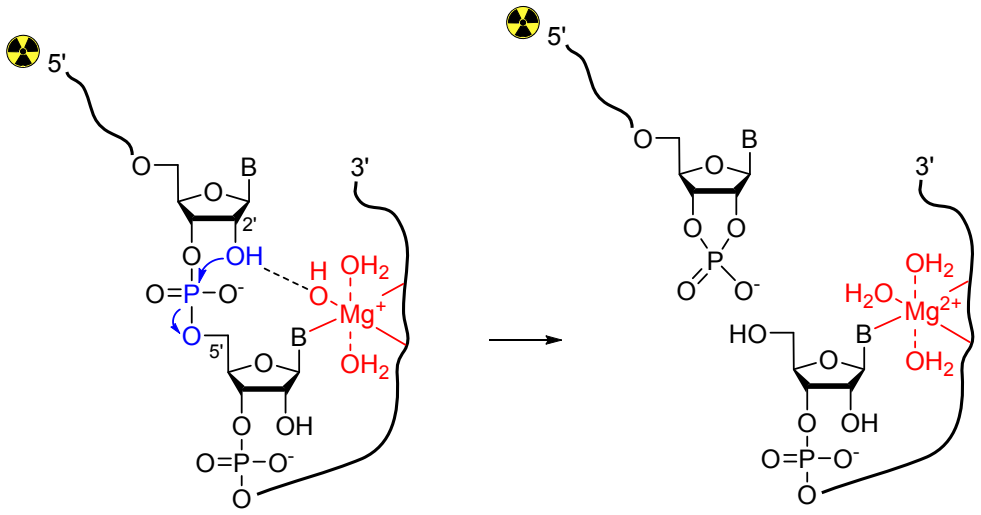


Figure 21: Self-cleavage of RNA promoted by the presence of metal ions. In-line nucleophilic attack of the 2'-OH with 5'-oxygen as leaving group cleaves the phosphodiester linkage (S_N2). The reaction is catalyzed by the magnesium-aqua-hydroxo-complex (red) via deprotonation of the nucleophile. For visualization purpose, the RNA is [^{32}P]-5'-end labeled (radioactive symbol). *Based on figure from Dr. Sofia Gallo.*[86]

The method is based on the structure-dependent specific degradation of RNA, either in presence of divalent metal ions or acid/base.[89] Single-stranded regions are flexible and therefore susceptible to self-cleavage, in contrast to highly structured regions of folded RNA. The dominant pathway of RNA degradation involves an 'in-line' conformation of the 2'-oxygen with the phosphorous and the adjacent 5'-oxygen. This allows for a nucleophilic attack of the 2'-oxygen at the phosphodiester backbone (Figure 21). The cleavage products of the RNA can be separated via denaturing PAGE and give a characteristic pattern which can then be compared in presence or absence of the metabolite and its derivatives. For a highly sensitive detection, [^{32}P]-5'-end labeled RNA is used. By means of in-line probing titration experiments the dissociation constant K_D can be determined, which represents the concentration of ligand required to convert half of the aptamer into its ligand-bound form.[49]

2.4 Aim of the Chapter

Since the discovery of the *btuB* riboswitch of *E. coli* as one of the first representatives of the gene-regulating RNA riboswitches, many efforts have been undertaken to elucidate their structure and function. Structural mapping, transcriptional pausing studies, and comparison to crystal structures of similar riboswitches revealed details on the *btuB* riboswitch conformation in absence and presence of the metabolite B_{12}. Further studies on the interaction of the riboswitch with different natural occurring or chemically modified B_{12} derivatives gave insights into the binding specificity. However, the binding of B_{12} to the riboswitch and the consequent structural switch have not been studied dynamically.

This chapter is focused on the dye labeling of the B_{12} metabolite to investigate the complex mechanism of B_{12} binding to the *btuB* riboswitch of *E. coli* and its structural rearrangement by use of Förster Resonance Energy Transfer. Compared to other methods like X-ray crystallography, FRET allows investigating dynamic folding processes of complex molecules or binding events in real time. FRET was already applied on the single-molecule level to study conformational changes of RNA upon binding of metals or metabolites.[42,50] So far the metabolites binding to the investigated riboswitches were rather small molecules such as *S*-adenosylmethionine (SAM) or thiamine pyrophosphate (TPP).[40,90] In the case of the *btuB* riboswitch the metabolite is B_{12}, which is a fairly large biomolecule. Cob(III)alamin itself is not fluorescent and can therefore not be visualized by fluorescence techniques. However, the structure of this metabolite gives the opportunity for fluorescence labeling, allowing to investigate its binding to the RNA, together with the induced RNA folding.

Herein, the synthesis and characterization of new dye-labeled B_{12}-derivates is described (Figure 22).

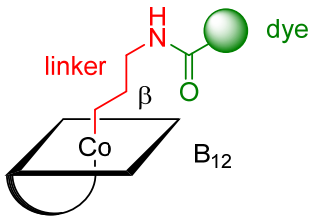


Figure 22: Schematic structure of dye-labeled B_{12}-derivatives. Attachment of a dye (green) to the β-face of cobalamin (black) with an aminopropyl linker (red).

The fluorescent dyes of choice are the common FRET pair Cy3 and Cy5, which are well established for nucleic acid investigations (Chapter 1.2.2). Alternatively, Cy5 is replaced by the ‘black hole quencher‘ BHQ-2 in order to minimize possible background issues that could arise due to fluorescence of the labeled metabolite in high concentrations. The donor dye Cy3 can be located at the B_{12} and the acceptor dye Cy5 or BHQ-2, respectively, at the RNA,

or vice versa. This reduces the possibility of misinterpreted artifacts in smFRET as both labeling schemes should give the same results. Furthermore, two derivatives of Cy3 and Cy5 differing in their substituents and charge were chosen to find the optimal label at the metabolite, that induces a correct riboswitch folding. Cy3 and Cy5 are positively charged, and BHQ-2 is uncharged. Sulfo-Cy3/Cy5 are negatively charged at the sulfonate groups and have thus a reduced possibility of intercalation into nucleic acids due to electrostatic repulsion (Chapter 1.2). For the attachment of the dye the axial position on the β-face of the cobalamin was chosen since this position of the cobalamin is known to influence only the binding affinity of the B_{12} to the riboswitch, but not the folding. Modifications at the corrin ring typically lead to a different folding of the RNA (see Chapter 2.3.4).[87] As the probes are designed to be used in folding investigations of the RNA, attachment at the side chains of the corrin ring would therefore not be suitable. Thus a linker group, which carries a suitable functionality, is introduced at the cobalt center of B_{12}, allowing for the coupling of the dyes.
An important assumption using this methodology is that the modified B_{12} derivatives do not change the riboswitch folding drastically. Therefore, interactions of the novel dye-labeled B_{12} derivatives with the *btuB* riboswitch are studied by in-line probing experiments.

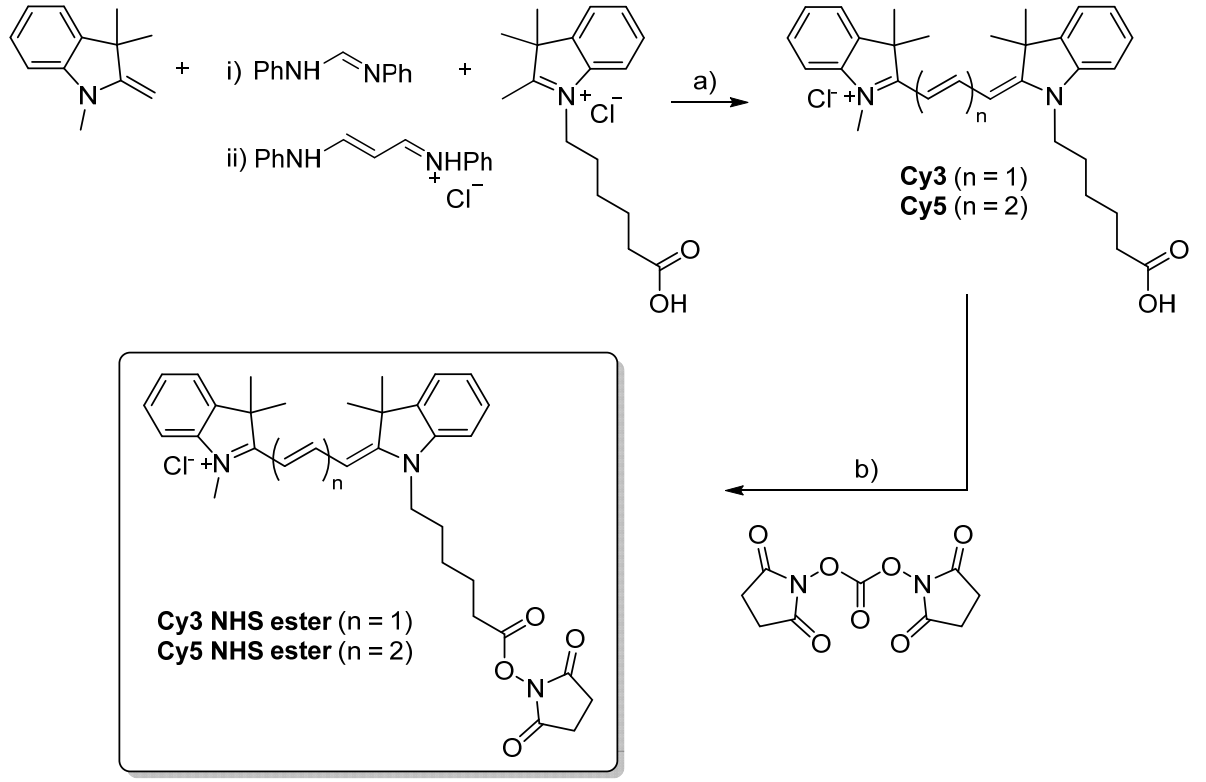
2.5 Results

2.5.1 Synthesis of Dye-labeled B_{12} Derivatives

The synthetic route towards dye-labeled B_{12} derivatives starts from hydroxocobalamin. In the first step, a primary amine function is attached to the cobalt center of the cobalamin on the β-face of the corrin ring. To this amine function the dyes are subsequently coupled via *N*-hydroxysuccinimide (NHS) ester chemistry. All reactions and analyses of B_{12} derivatives and dyes were done in the dark to prevent from decomposition or deactivation.

Cyanine Dyes Cy3, Cy5 and NHS Esters

Commercially available cyanine dyes are rather expensive, especially in their active forms such as azides, phosphoramidites and NHS esters. Therefore **Cy3** and **Cy5** were synthesized in a one-pot reaction approach (Scheme 2).



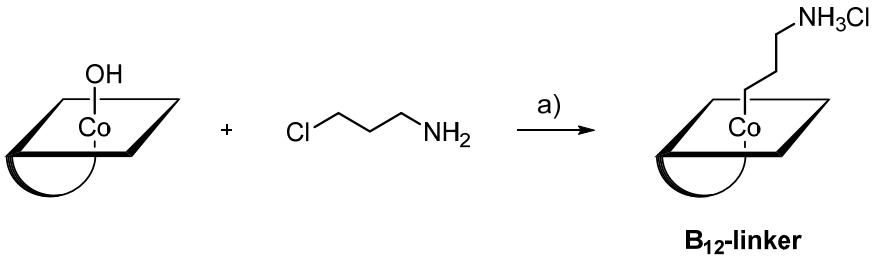
Scheme 2: Synthesis of Cy3 and Cy5 NHS esters. Synthesis of (i) Cy3 NHS ester and (ii) Cy5 NHS ester starting from 2-methylene-1,3,3-trimethylindoline, 1-(5-carboxypentyl)-2,3,3-trimethyl-3*H*-indolium chloride, (i) *N,N'*-diphenylformamidine, and (ii) *N*-((1*E*)-3-(phenylimino)prop-1-enyl)benzenamine hydrochloride via **Cy3**/**Cy5**, followed by addition of DSC; a) NaOAc, EtOH, 24 h, 85 °C, (i) yield: 17 %; ii) yield not determined; b) DIPEA, DMF, 2 h, 60 °C.

Indole, *N,N'*-diphenylformamidine or malondialdehyde dianil hydrochloride, and the carboxypentyl indolium salt were brought to reaction to obtain the cyanine acid **Cy3** or **Cy5**.[91,92] Due to side product formation the yield was low compared to the multi-step approach.[93]

NHS esters are sensitive to hydrolysis. To avoid this deactivation during long storage, **Cy3** or **Cy5 NHS esters** were prepared *in situ* by addition of DSC to the Cy acid. These derivatives were used for the following coupling to the respective B_{12} derivative.

B_{12}-linker

The addition of a linker to B_{12} providing a suitable functional group gives the possibility to attach various dyes to the B_{12}. This function is fulfilled by a primary amine group to which fluorophores can be attached via activated esters, forming a stable amide bond. **B_{12}-linker** was synthesized according to the literature procedure.[94] Starting from commercially available hydroxocobalamin (OHCbl), 3-aminopropylchloride was attached to the cobalt center of cobalamin replacing the hydroxo moiety in the upper β-face position (Scheme 3). Zinc dust was used as a reducing agent. The cobalt center in hydroxocob(III)alamin is reduced to Co^{+}, followed by the alkylation via bimolecular nucleophilic substitution (S_N2) to give β-(3-aminopropyl)cobalamin as a Co^{3+} species.[95] As mentioned before, the synthesis was carried out in the dark, as the newly formed organometallic cobalt-carbon bond is sensitive to cleavage by light.

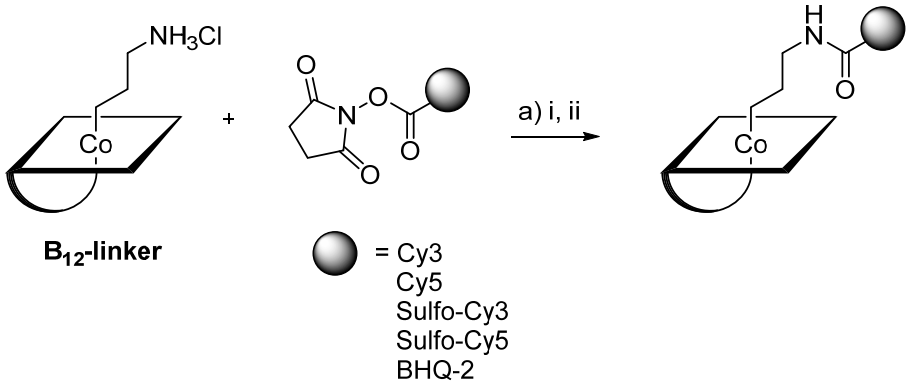


B_{12}-linker

Scheme 3. Synthesis of B_{12}-linker. Reaction of hydroxocobalamin and 3-aminopropylchloride; a) Zn-dust, NH_4Cl, H_2O, 5 h, r.t., yield: 82 %.

B_{12}-dye

The fluorescent cyanine dyes Cy3, Cy5, Sulfo-Cy3 and Sulfo-Cy5, as well as the 'black hole quencher' BHQ-2 were used as NHS esters and coupled to the primary amine group of the linker in **B_{12}-linker**. The reactions yielding the B_{12}-Cy derivatives **B_{12}-Cy3**, **B_{12}-Cy5**, **B_{12}-Sulfo-Cy3** and **B_{12}-Sulfo-Cy5** were carried out at room temperature in the presence of DIPEA and under nitrogen atmosphere to avoid hydrolysis of the NHS ester (Scheme 4).



Scheme 4. Synthesis of B_{12}-dye derivatives. Synthesis of B_{12}-dye derivatives starting from **B_{12}-linker** and dye NHS ester; a) i: DIPEA, DMF, 24 h, r.t., yields: 40-70 % (Cy3, Cy5, Sulfo-Cy3, Sulfo-Cy5); ii: 0.1 M CO_3^{2-}/HCO_3^- buffer (pH 8.3), r.t., yield: 48 % (BHQ-2).

The synthesis of **B_{12}-BHQ-2** was carried out by reacting **B_{12}-linker** with BHQ-2 NHS ester in carbonate-bicarbonate buffer at pH 8.3. The buffer and pH were chosen to be optimal for the NHS ester crosslinking reaction.

B_{12}-dye products were obtained in yields of 40-70 %. Each derivative was characterized by UV/Vis, HPLC, NMR and ESI-MS/HR-ESI-MS (Chapter 2.5.2).

2.5.2 Characterization of Dye-labeled B_{12} Derivatives

UV/Vis Spectroscopy

UV/Vis spectra of the B_{12}-dye derivatives in water show the characteristic absorption bands of cobalamin combined with the ones corresponding to the respective dye. **B_{12}-linker** reveals the main peaks at 266, 281, 317, 349, 375 and 526 nm (Figure 23), which are in agreement with the literature values of unmodified AdoCbl (260, 288, 315, 340, 375, 522 nm, see Appendix).[57] In spectra of **B_{12}-Cy3** and **B_{12}-Sulfo-Cy3** (Figure 23 a and Appendix) the characteristic absorption peaks for the Cy3 dye are observed at λ_{max} = 545 nm (Cy3) and λ_{max} = 551 nm (Sulfo-Cy3), together with a shoulder at 514 nm (Cy3) and 518 nm (Sulfo-Cy3), respectively. Likewise, in spectra of **B_{12}-Cy5** and **B_{12}-Sulfo-Cy5** (Appendix) the characteristic absorption peaks for the Cy5 dye are observed at λ_{max} = 642 nm (Cy5) and λ_{max} = 648 nm (Sulfo-Cy5), as well as a shoulder at 599 nm (Cy5) and 604 nm (Sulfo-Cy5), respectively. **B_{12}-BHQ-2** shows the absorption maximum of the BHQ-2 at 559 nm (Figure 23 b).

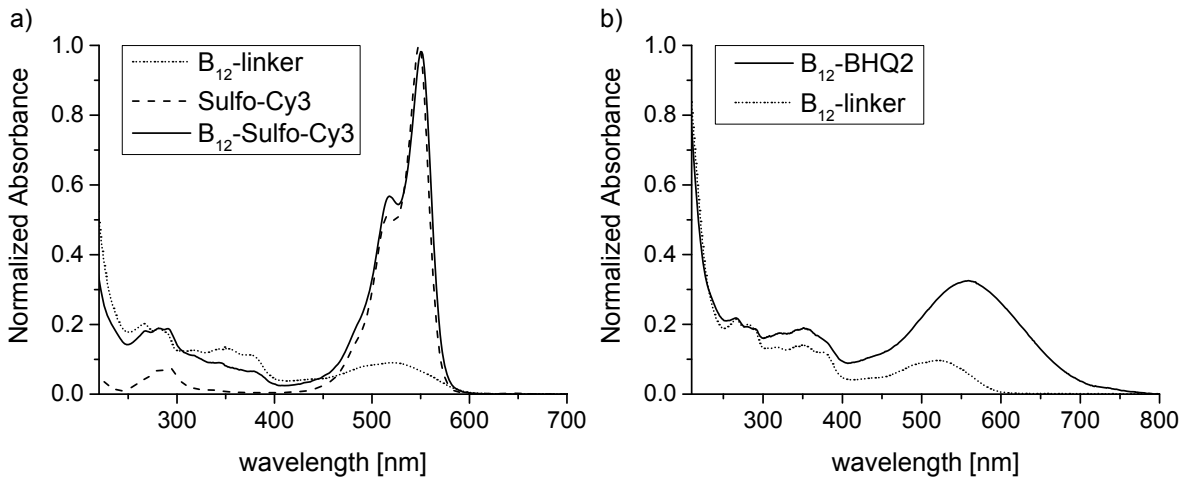


Figure 23: UV/Vis spectra of B_{12} derivatives. Normalized absorbance spectra in water at r.t. of a) **B_{12}-linker** (dotted line), Sulfo-Cy3 (dashed line) and **B_{12}-Sulfo-Cy3** (solid line), b) **B_{12}-linker** (dotted line) and **B_{12}-BHQ-2** (solid line).

HPLC

HPLC analysis of **B_{12}-linker** shows an increased retention in comparison to hydroxocobalamin (OHCbl) due to the replacement of the hydroxyl moiety by the less polar aminopropyl linker (Figure 24).

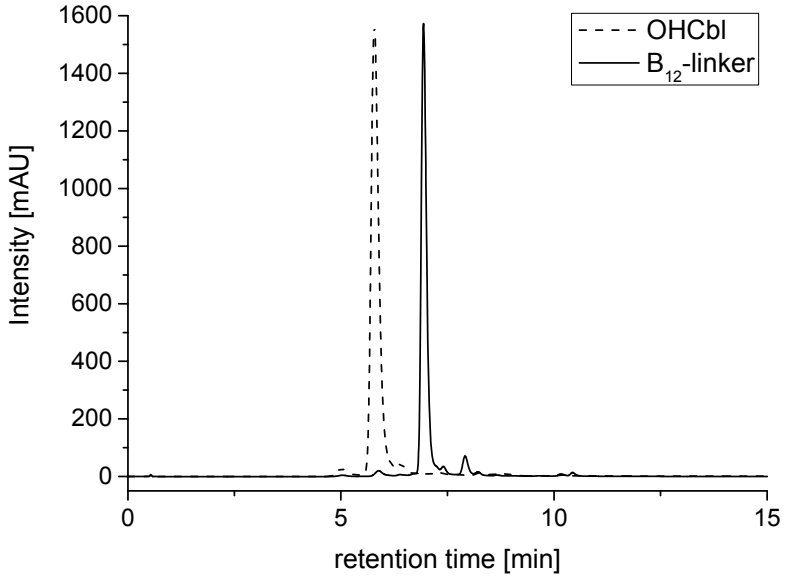


Figure 24: HPLC chromatograms of B_{12}-linker and OHCbl recorded at 360 nm. OHCbl (dashed line) elutes 1 min earlier than **B_{12}-linker**.

B_{12}-dye derivatives were identified by HPLC analysis with product peaks at a retention time of around 15 min for Cy3/Cy5 derivatives, 11 min for sulfonated Cy3/Cy5 derivatives, and 18 min for the BHQ-2 derivative under the applied conditions.

A characteristic comparison of HPLC traces of starting material (**B_{12}-linker**) and product is shown in an overlay chromatogram (Figure 25). The clear separation of **B_{12}-linker** at t_R = 7 min from the product **B_{12}-Cy5** is observed. The signal overlay of the HPLC traces for **B_{12}-Cy5** at 360 nm, 650 nm and fluorescence emission at 670 nm shows the absorption and emission of the individual components B_{12} and Cy5.

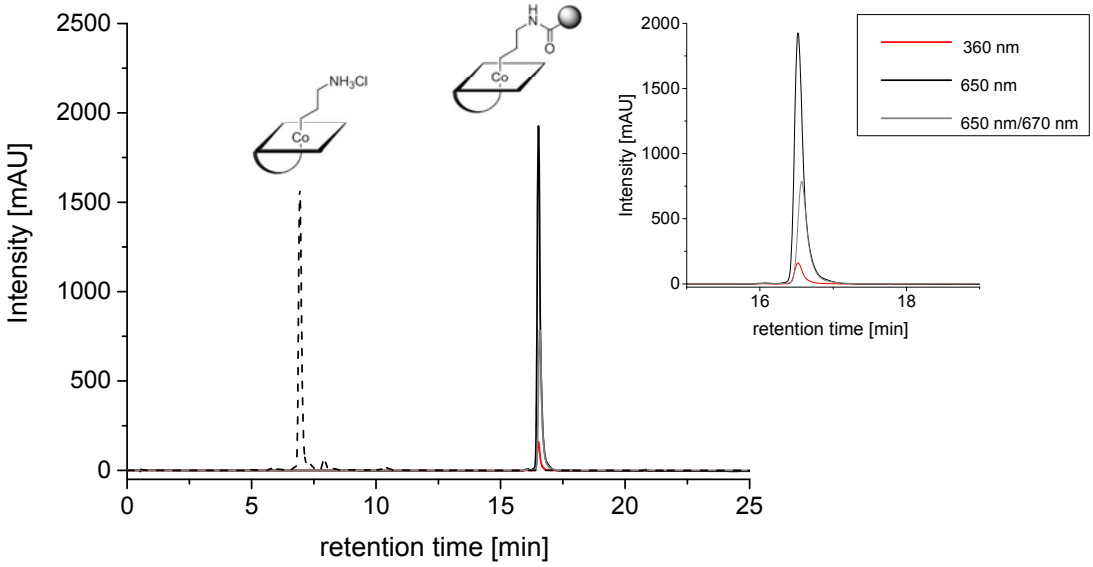


Figure 25: HPLC chromatograms of B_{12}-linker and B_{12}-Cy5. B_{12}-linker (dashed line) with a retention time of 7 min is clearly separated from **B_{12}-Cy5** (solid line) at 16.5 min. The successful attachment of Cy5 to B_{12} in **B_{12}-Cy5** can be observed in the overlay of chromatograms at the different wavelength for the components 360 nm (Cbl), 650 nm (Cy5) and fluorescence emission at 670 nm (Cy5, excitation at 650 nm). The inset shows a close-up of the region 15 to 19 min.

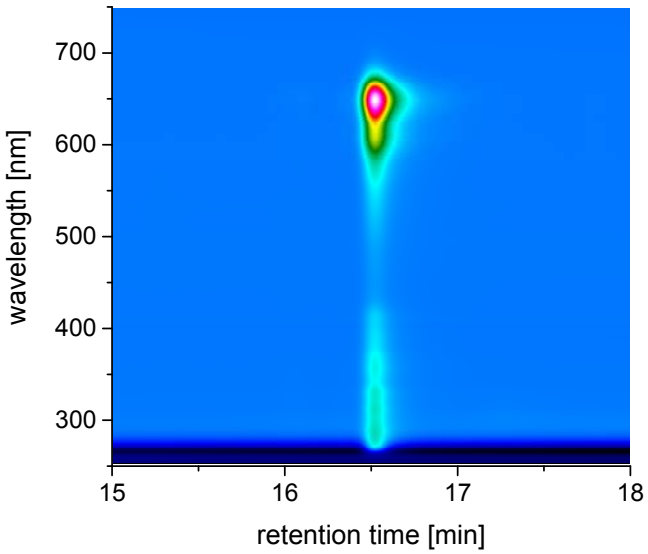


Figure 26: Contour plot of B_{12}-Cy5. 3D field of the HPLC DAD. The maximum absorption is detected at 650 nm (Cy5). The characteristic cobalamin absorption is detected at around 300 nm to 360 nm.

Additionally, the wavelengths of the different functionalities can be seen in the 3D field contour plot of **B_{12}-Cy5** measured with the diode array detector (DAD) in an HPLC experiment (Figure 26). The same is observed for the other derivatives (data not shown).

NMR Spectroscopy

For the characterization of the novel B_{12} derivatives ^{1}H- and ^{13}C{^{1}H}-NMR experiments were performed. Most of the peaks could be assigned with these and the 2D experiments [^{1}H,^{1}H]-COSY (correlation spectroscopy), [^{1}H,^{13}C]-HSQC (Heteronuclear Single Quantum Coherence) and [^{1}H,^{13}C]-HMBC (Heteronuclear Multiple Bond Correlation). HSQC experiments show the correlation between chemical shifts of directly bound nuclei, and HMBC of nuclei separated by two or more chemical bonds. The atom numbering of the different compounds is depicted in Figure 27.

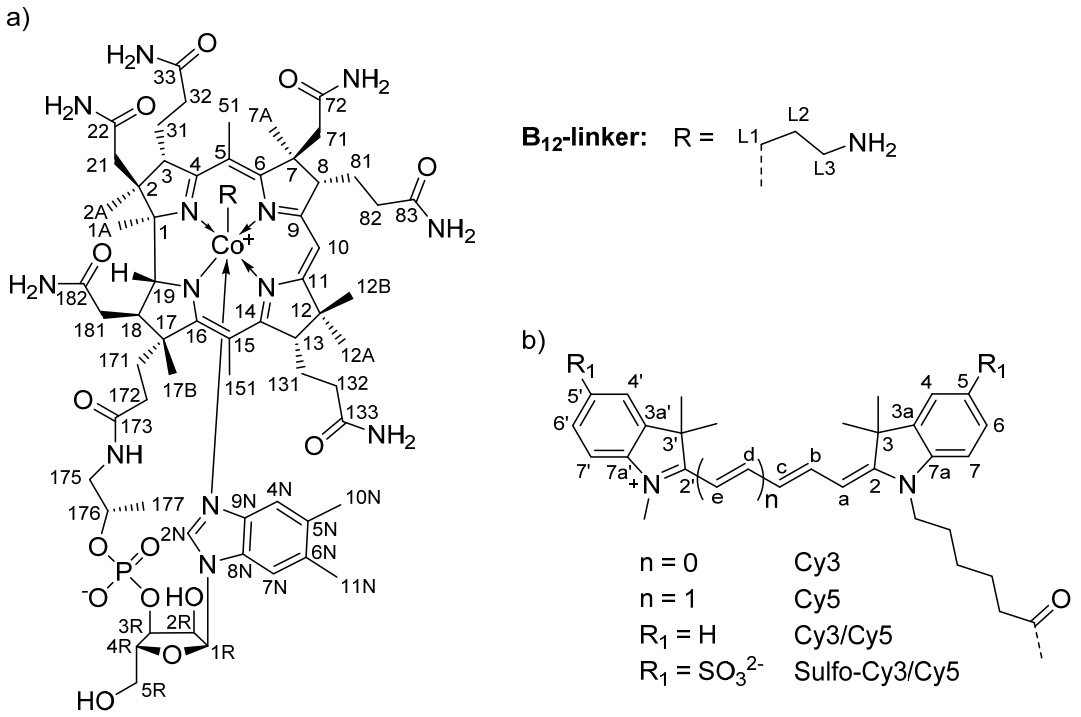


Figure 27: Nomenclature for the atom positions of B_{12}, B_{12}-linker, and cyanine dyes. Atom numbering of a) B_{12}, B_{12}-linker, and b) the cyanine dyes Cy3, Cy5, Sulfo-Cy3 and Sulfo-Cy5.

In the ^{1}H-NMR spectrum of **B_{12}-linker** the characteristic peaks for B_{12} are observed (Figure 28). In the aromatic region between 7.5 and 6 ppm the peaks of 2N, 4N and 7N of the benzimidazole moiety, 1R of the ribose, and 10 of the corrin ring can be seen. The assignment is in perfect agreement with the literature.[96] Another significant region is located between 5 and 3 ppm. Here, 2R, 3R, 4R and 5R of the ribose, 3, 8, and 19 of the corrin ring, and 175, 176 are found (see inset in Figure 28). Additionally, the peaks for the newly attached linker can be seen in the high field region between -0.5 and 2.7 ppm. Notably, protons closest to the cobalt center have a negative shift.

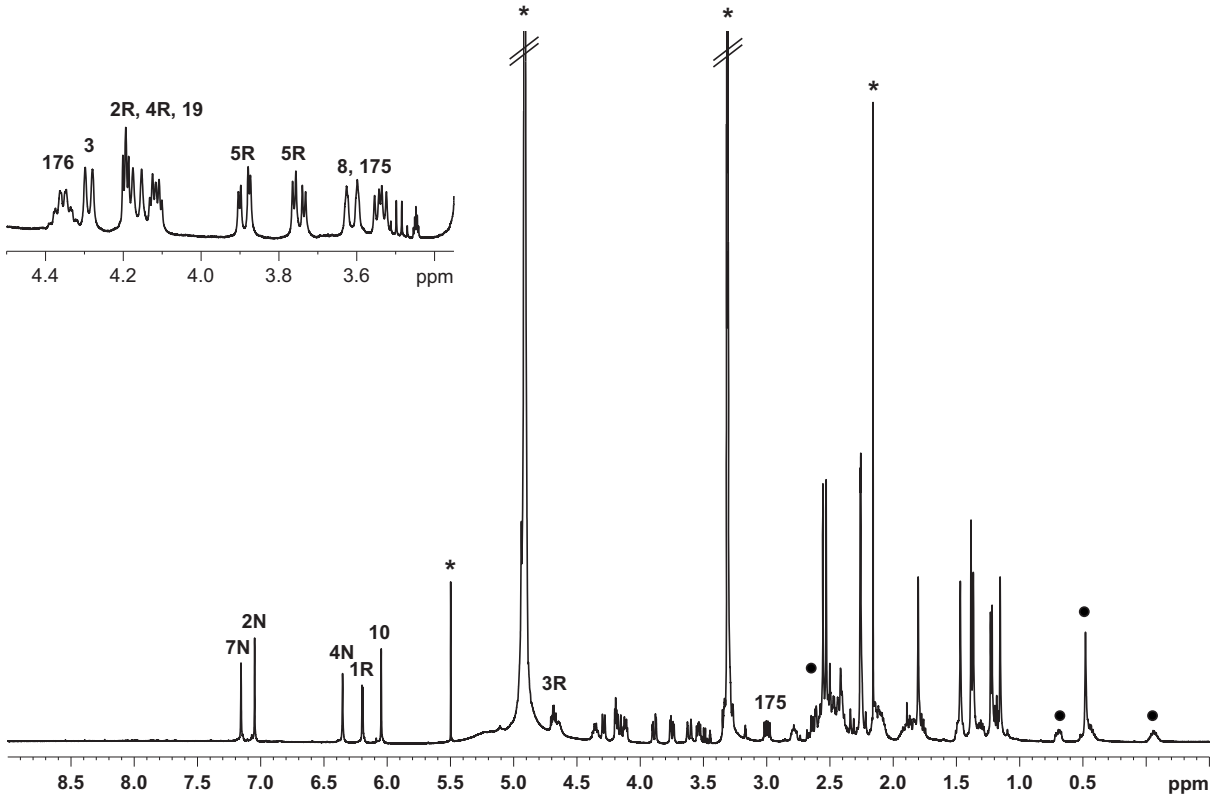


Figure 28: ^{1}H-NMR spectrum of B_{12}-linker in MeOD-d_4. The signals of the most characteristic protons are indicated. Protons belonging to the attached linker group are marked with black dots (●). The inset shows a close-up of the region 4.5 to 3.5 ppm.

The successful coupling of **Cy3** to **B_{12}-linker** is observed in the NMR spectrum of **B_{12}-Cy3** (Figure 29). In the aromatic region, the signals of B_{12} (2N, 4N, 7N, 1R and 10), the signals of the alkene chain in Cy3 (a, b and c), and of the Cy3 phenyl rings (4, 5, 6 and 7) can be assigned. Protons of the carboxypentyl chain are located high field between 2.3 and 1.5 ppm.

Analog to **B_{12}-Cy3** the corresponding signals are observed for **B_{12}-Cy5** with the additional CH_2 signals at 8.3 ppm and 6.3 ppm, and a small shift of a, b and c of the alkene chain (Figure 30).

The sulfonated derivatives **B_{12}-Sulfo-Cy3** and **B_{12}-Sulfo-Cy5** reveal a shift of the phenyl signals 4 and 6 to down field due to the electron-withdrawing sulfonate groups at C5 (Appendix).

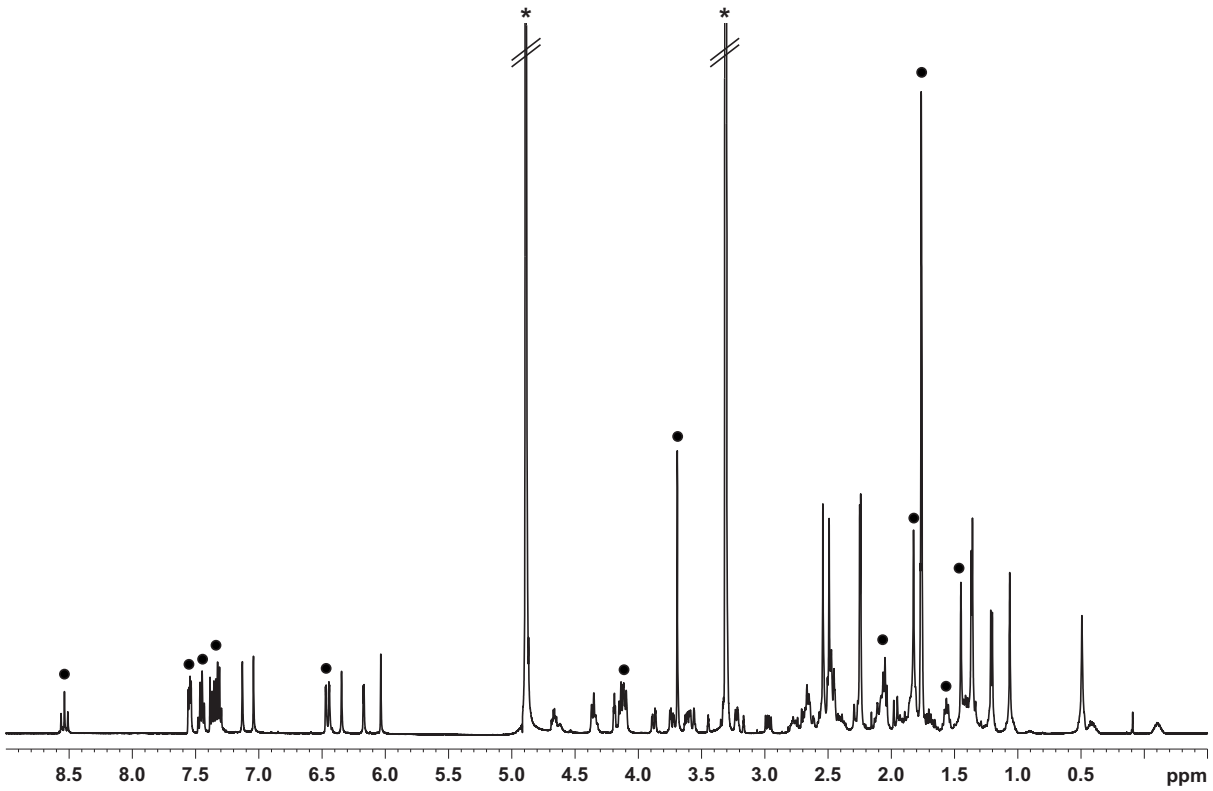


Figure 29: ^{1}H-NMR spectrum of B_{12}-Cy3 in MeOD-d_4. Proton signals belonging to Cy3 are marked with black dots (●). Other signals are assigned to B_{12}.

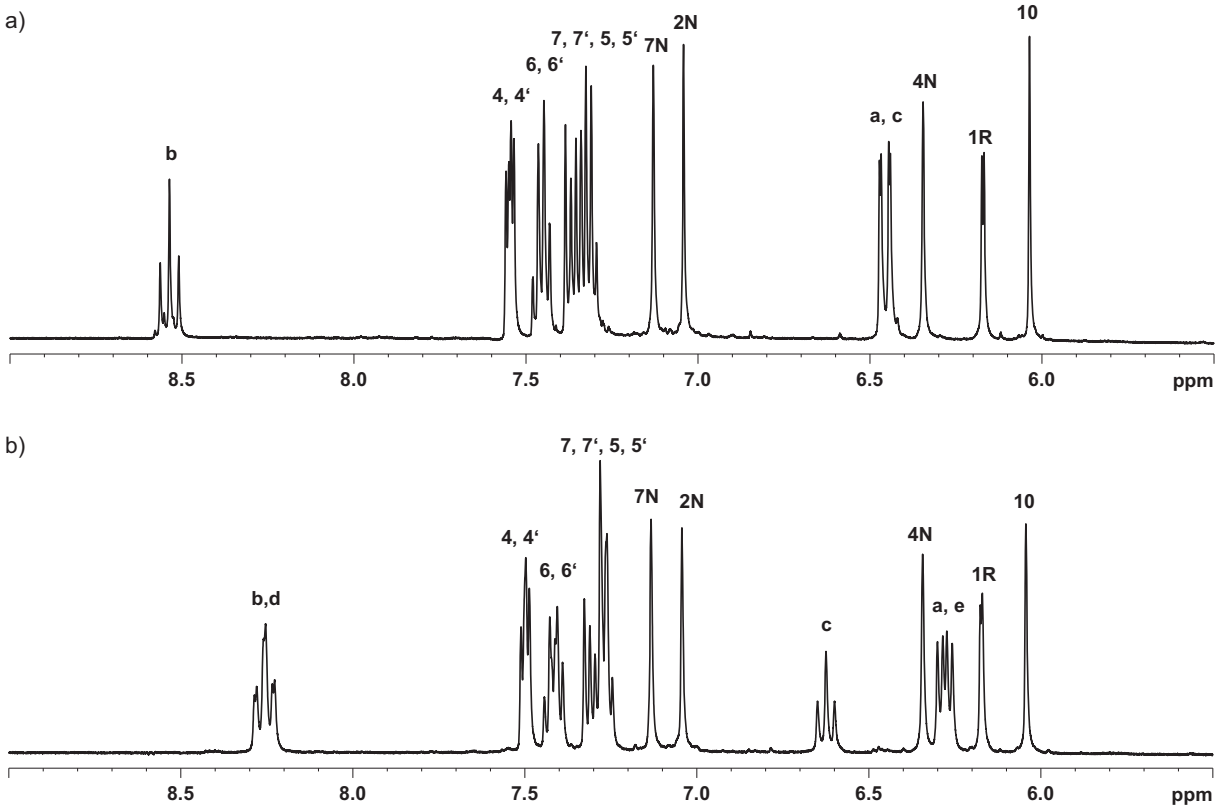


Figure 30: ^{1}H-NMR spectra of B_{12}-Cy3 and B_{12}-Cy5 in MeOD-d_4. Aromatic region of a) **B_{12}-Cy3** and b) **B_{12}-Cy5**.

The designated atom numbering of **B_{12}-BHQ-2** is shown in Figure 31.

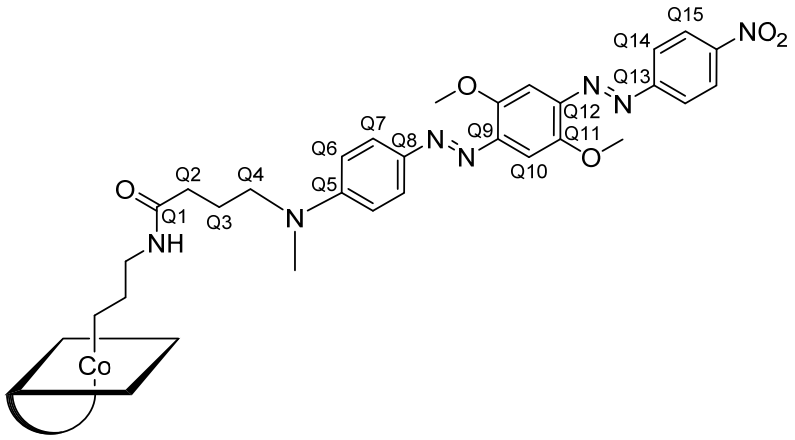


Figure 31: Nomenclature for the atom positions of BHQ-2 in B_{12}-BHQ-2.

In the NMR spectrum of **B_{12}-BHQ-2** (Figure 32), the signals in the aromatic region can be assigned to the protons Q6, Q7, Q10, Q14 and Q15 of the BHQ-2 phenyl rings. The methoxy groups show peaks at 4.0 ppm and NMe at 3.1 ppm. Alkyl proton signals Q2 and Q3 are located high field at 2.1 and 1.8 ppm, and Q4 at 3.5 ppm.

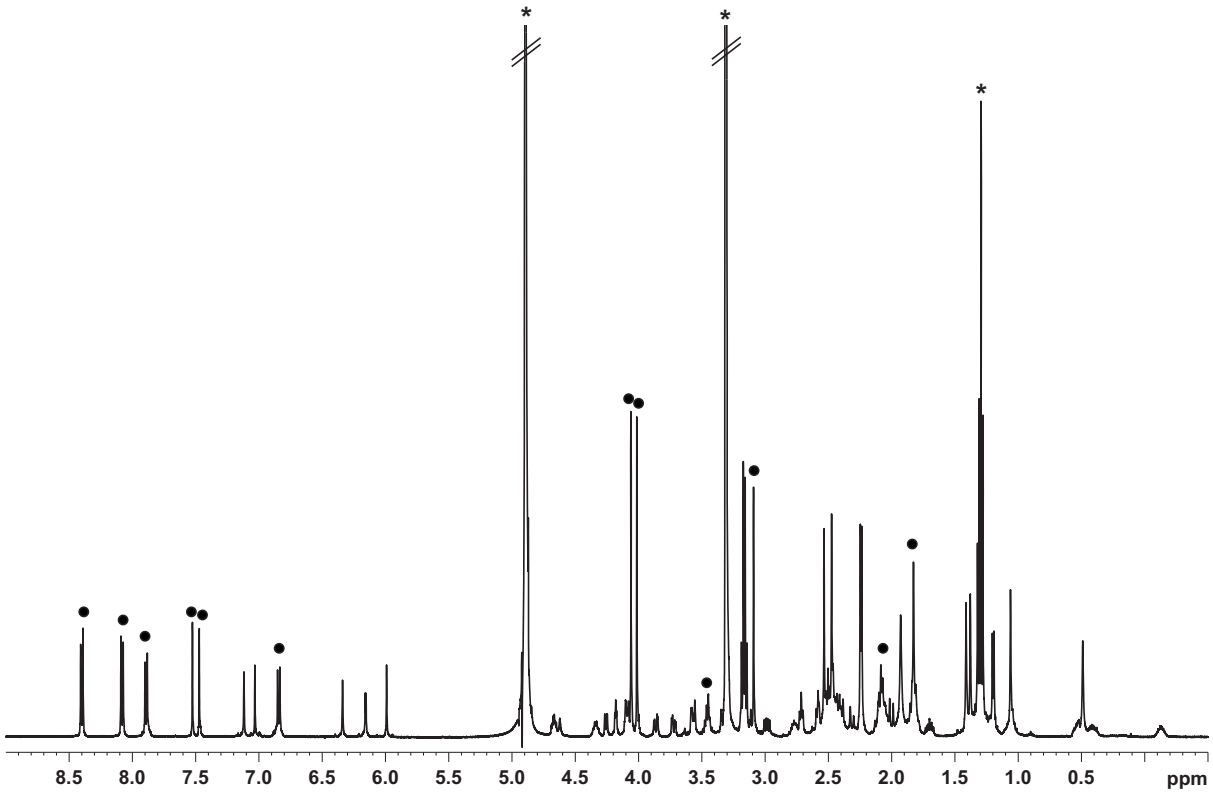


Figure 32: ^{1}H-NMR spectrum of B_{12}-BHQ-2 in MeOD-d_4. Proton signals belonging to BHQ-2 are marked with black dots (●). Other signals are assigned to B_{12}.

Mass Spectrometry

In ESI-MS and high resolution (HR)-ESI-MS measurements of the B_{12} derivatives **B_{12}-linker**, **B_{12}-Cy3**, **B_{12}-Cy5**, **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** and **B_{12}-BHQ-2** the respective products peaks were detected. Furthermore, fragments due to Co-C bond cleavage during the measurement are observed. The **B_{12}-linker** product peak ($[M]^+$ calcd.: 1387.6 m/z) is detected in the ESI-MS positive mode at 1387.7 m/z, as well as the protonated doubly charged compound at 694.4 m/z. In HR-ESI-MS experiments of **B_{12}-Cy3** (Figure 33) all peaks are assigned to the desired product or to various product fragments present as sodium, methanol or protonated adducts. Main peaks are detected at m/z = 940.94318, being a doubly charged methanol-sodium adduct of the target compound **B_{12}-Cy3** ($[M+Na+MeOH]^{2+}$ calcd. for $C_{96}H_{135}CoN_{16}O_{16}PNa^{2+}$: 940.96159 m/z). The cobalamin fragment without the apical ligand is detected as singly charged species at 1351 m/z and doubly charged at 676 m/z with various adducts in the same region. The Cy3-linker fragment is located at m/z = 530 and adducts between m/z = 486 and 544.

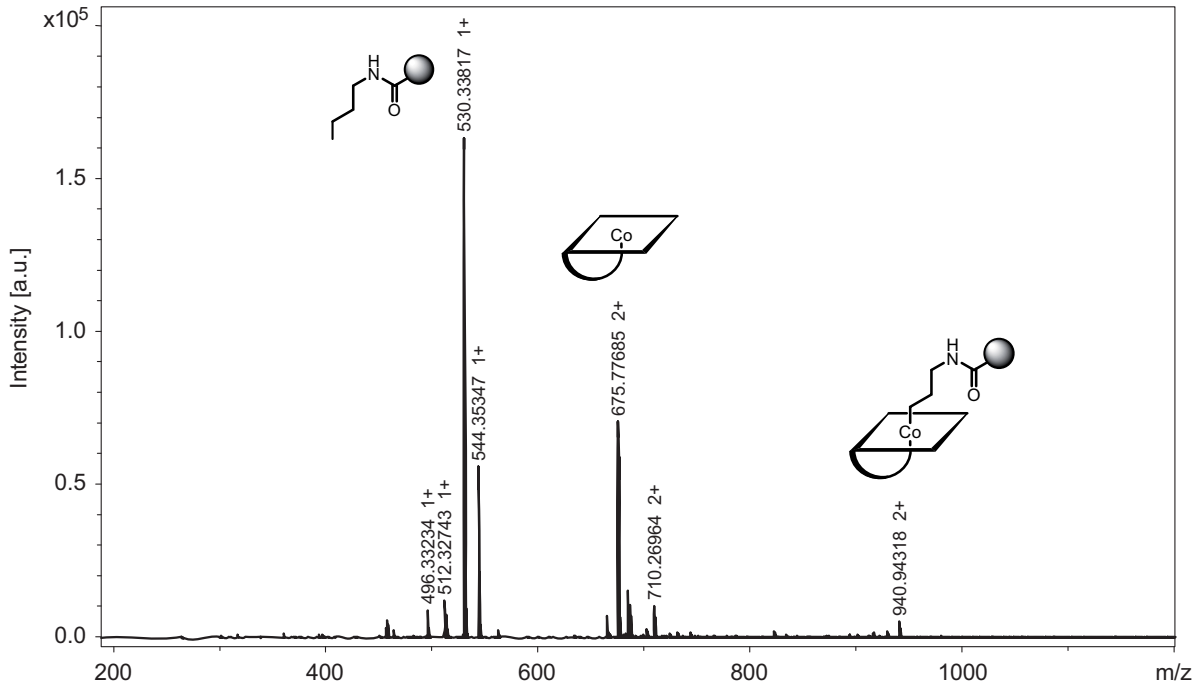


Figure 33: High resolution ESI-MS spectrum of B_{12}-Cy3. The product peak is located at m/z = 940 and the cleavage fragments at m/z = 530 and 674.

In positive mode ESI-MS measurements **B_{12}-Cy5** is detected at 1852.9 m/z ($[M]^+$) and 926.9 m/z ($[M+H]^{2+}$). In HR-ESI-MS, signals are observed at 942.96065 m/z, belonging to the doubly charged protonated methanol adduct of **B_{12}-Cy5** ($[M+H+MeOH]^{2+}$), and at 953.95183 m/z, belonging to the doubly charged methanol-sodium adduct ($[M+Na+MeOH]^{2+}$ calcd. for $C_{98}H_{137}CoN_{16}O_{16}PNa^{2+}$: 953.96942 m/z).

The sulfonated B_{12} derivatives are detected in HR-ESI-MS measurements (negative mode) with the **B_{12}-Sulfo-Cy3** main peak at 991.89869 ($[M-H]^{2-}$ calcd. for $C_{95}H_{128}CoN_{16}O_{21}PS^{2-}$: 991.89978 m/z), and **$B_{12}$-Sulfo-Cy5** at 1004.90741 m/z ($[M-H]^{2-}$ calcd. for $C_{97}H_{130}CoN_{16}O_{21}PS^{2-}$: 1004.90760 m/z).

In ESI-MS measurements the product peak of **B_{12}-BHQ-2** is observed at m/z = 1875.8 ($[M+H]^{+}$) and its sodium adduct at 1898.7 m/z ($[M+Na]^{+}$).

Quantum Yields and Steady-State Anisotropy

The fluorescence quantum yield of a dye is an important property in FRET experiments (Chapter 1.2.1). Cobalamin has been reported to quench the fluorescence of covalently bound fluorescein derivatives due to the overlap of the electronic orbital of cobalamin with the excited state of the dye.[94] The fluorescence quantum yields (Φ_s) of the novel B_{12}-cyanine conjugates were therefore determined and compared to the free cyanine dyes Cy3 and Cy5 (Table 1). Indeed, the quantum yields of the B_{12} derivatives are reduced. **B_{12}-Cy3** revealed a quantum yield that is 6.5-fold lower, and **B_{12}-Sullfo-Cy3** a quantum yield that is 10-fold lower than their corresponding free dyes. The quantum yield of Cy5 is 2.5-fold and of Sulfo-Cy5 4-fold higher, compared to their corresponding B_{12} derivatives (Figure 34). **B_{12}-Cy5** and **B_{12}-Sulfo-Cy5** are thus from this point of view more suitable for FRET experiments than the other derivatives. Cy3 or Sulfo-Cy3 can then be placed at the RNA to complete the FRET pair.

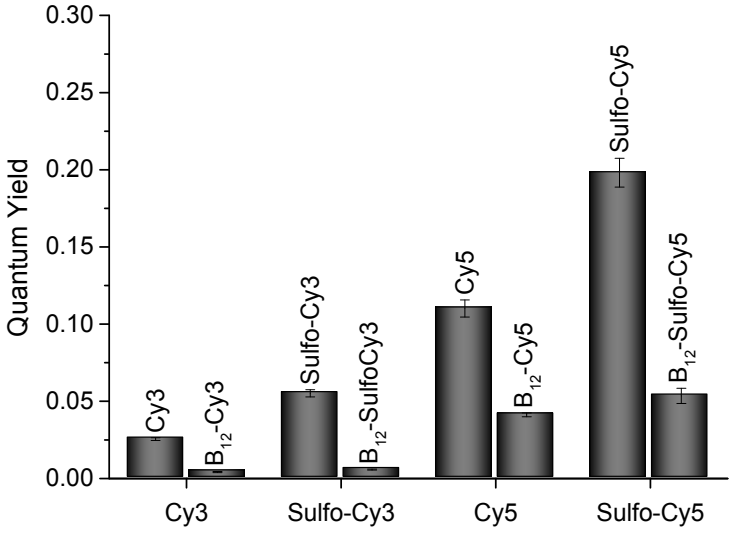


Figure 34: Quantum yields of free dyes and B_{12}-dye derivatives in *PBS buffer*. The quantum yields of the B_{12} derivatives are reduced compared to the corresponding free dyes.

Additionally, steady-state fluorescence anisotropy (r) was measured to determine the changes of rotational freedom of the dye when attached to B_{12} (Table 1). No differences were observed between unbound Cy3 and B_{12}-Cy3 conjugates with all values being around

r = 0.25. B_{12}-Cy5 derivatives exhibit a value of 0.26 (0.23 sulfonated), while free Cy5 has a lower anisotropy value of 0.15 (0.13 sulfonated).

Table 1. Fluorescence quantum yields (Φ_s) and static fluorescence anisotropy values (r) of free Cy3 or Cy5 derivatives and their B_{12} conjugates in *PBS buffer* (pH 7.4) at 22 °C.

Compound	Φ_s	r
Cy3	0.026 ± 0.001[a]	0.263 ± 0.006
Sulfo-Cy3	0.055 ± 0.002[a]	0.247 ± 0.005
B_{12}-Cy3	0.004 ± 0.0004	0.255 ± 0.008
B_{12}-Sulfo-Cy3	0.006 ± 0.0003	0.247 ± 0.009
Cy5	0.110 ± 0.006[b]	0.153 ± 0.008
Sulfo-Cy5	0.198 ± 0.009[b]	0.133 ± 0.005
B_{12}-Cy5	0.041 ± 0.001	0.258 ± 0.010
B_{12}-Sulfo-Cy5	0.054 ± 0.005	0.225 ± 0.012

[a] Cy3 derivative in PBS Φ = 0.04[23]; [b] Cy5 derivative in PBS Φ = 0.20[23]

Photostability

As mentioned earlier (Chapter 2.2.1), the Co-C bond between the cobalamin and the linker can be cleaved by photolysis. The formed B_{12} radical can further react to hydroxocobalamin. To verify the stability of the B_{12}-dye compounds, absorbance and fluorescence spectra exemplary of **B_{12}-Cy3** were recorded before and after irradiation for 10 min at 532 nm, which corresponds to the standard measuring time and the wavelength of the green laser used in FRET experiments (Figure 35).

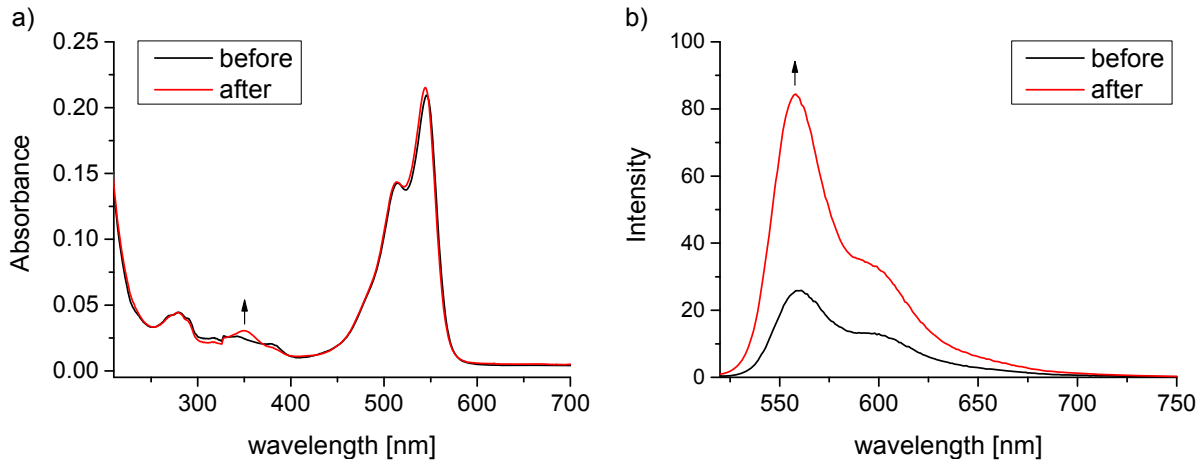


Figure 35: Absorbance and fluorescence spectra of B_{12}-Cy3. a) Absorbance spectra and b) fluorescence spectra of **B_{12}-Cy3** before (black) and after (red) irradiation at 532 nm for 10 min.

After irradiation, the absorbance peak of the cleavage product hydroxocobalamin appears at 350 nm, and at the same time the fluorescence intensity is increased. This shows once more the quenching effect of B_{12}, which is reduced upon cleavage of Cy3 from B_{12}.

2.5.3 Interaction of the *btuB* Riboswitch and Dye-labeled B_{12} Derivatives

Modifications of AdoCbl, the natural metabolite of the *btuB* riboswitch of *E. coli*, can lead to an altered structural conformation of the RNA and a changed binding behavior of the B_{12} (Chapter 2.3.4). In-line probing was used to monitor the structural rearrangement of the *btuB* riboswitch upon binding of the modified B_{12} derivatives **B_{12}-linker**, **B_{12}-Cy3**, **B_{12}-Cy5**, **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** and **B_{12}-BHQ-2** in comparison to the natural ligand AdoCbl. In-line probing titration experiments were performed with the [^{32}P]-5'-end-labeled full-length aptamer sequence of the *btuB* riboswitch of *E. coli* designed by Gallo *et al*.[87] The RNA was therefore incubated in the presence of increasing concentrations of the modified B_{12} derivatives. The experiments were carried out in the dark to prevent photolysis of the Co-C bond. The cleavage products were separated by denaturing polyacrylamide gel electrophoresis (PAGE) and analyzed. The concentration of B_{12} was plotted versus the intensities of characteristic bands and fitted to a 1:1 binding isotherm[97] to obtain the respective binding affinity constants.

In presence of the natural ligand AdoCbl nine characteristic cleavage sites (Figure 36, marked with grey triangles) are observed in the *btuB* riboswitch. The resulting cleavage pattern was used for comparison in presence of the modified B_{12} derivatives. Upon addition of AdoCbl (+) the intensities of bands 1, 2b, 3, 4 and 8 decrease, whereas 2, 5, 6, and 7 increase compared to the negative control (-, incubation of RNA in absence of metabolite). Additionally to the previously described cleavage sites, cleavage site B (U19) increases with higher concentrations of AdoCbl (Figure 36, green triangle). Each experiment was carried out together with the described +/- controls to ensure a correct folding behavior of the riboswitch under the applied conditions. An additional control (c), which shows the unreacted riboswitch in absence of metal ions and cobalamin, was applied to ensure the purity of the labeled RNA. An OH ladder (alkaline hydrolysis), which is cleaving after every nucleotide, and a T1 ladder (RNAse T1 digestion), cleaving after each guanine residue, were used to assign the cleavage sites.

The fairly large dyes Cy3 and Cy5 were attached to the B_{12} (Chapter 2.5.1) to fluorescently label the metabolite, *e.g.* for FRET experiments. To rule out that the dyes themselves influence the riboswitch, control experiments with Cy3, Cy5, Sulfo-Cy3 and Sulfo-Cy5 were carried out. The incubation of the RNA with high concentrations of Cy3 or Cy5 carrying a positive charge prevented the RNA from entering the gel, possibly due to electrostatic interactions. The RNA is mainly detected in the upper part of the gel and in the full length

band indicating a suppressed cleavage reaction. In comparison, the incubation of the RNA with Sulfo-Cy3 or Sulfo-Cy5 (Figure 36) carrying negative charges at the sulfonate groups does not change the cleavage pattern, thus proving that the dyes do not interact with the riboswitch.

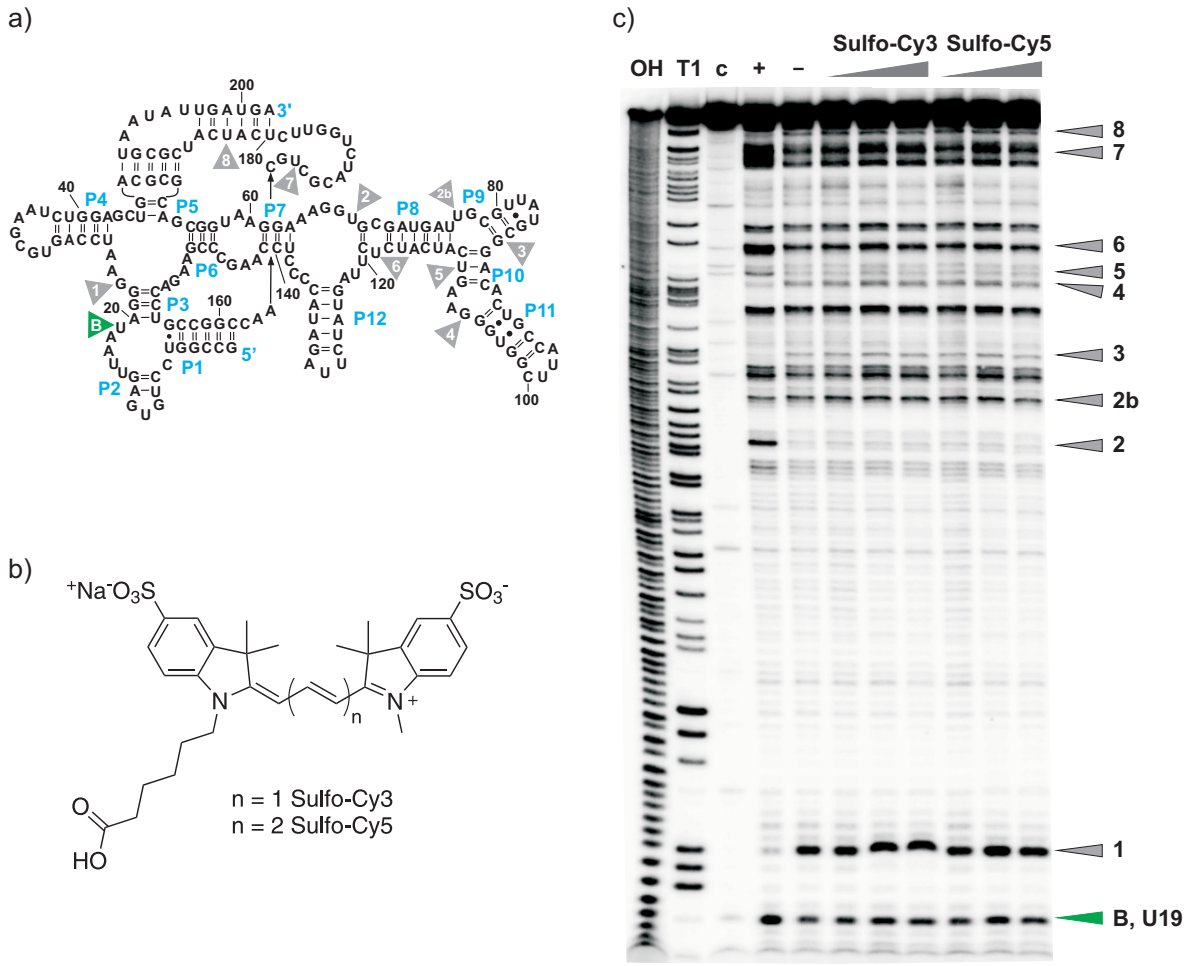


Figure 36: In-line probing experiments of the 202 nt long *btuB* riboswitch (pSG1) with Sulfo-Cy3 and Sulfo-Cy5. a) Secondary structure of pSG1 with the nine characteristic cleavage sites (marked in grey) and an additional cleavage site B (green). b) Structure of Sulfo-Cy3 and Sulfo-Cy5. c) Incubation of pSG1 in absence (–) and presence (+) of AdoCbl, and increasing concentrations of Sulfo-Cy3 and Sulfo-Cy5. Lane T1: RNAse T1 digestion; lane OH: alkaline hydrolysis ladder; lane c: control (unreacted RNA).

To compare the influence of the replacement of the upper ligand (adenosyl moiety) by the amine linker, the RNA was incubated with **B_{12}-linker**. However, no significant change in the intensity of the cleavage bands upon increasing B_{12} concentrations was observed, indicating that no structural switch of the RNA is induced.

Incubation of the *btuB* riboswitch with **B_{12}-BHQ-2**, which is an uncharged derivative, shows the same trends in all cleavage sites (Figure 37 a) as with AdoCbl (*vide supra*). Further cleavage sites situated at U58 (F) and A64 (G), as well as a, I, J, L, and M are observed. An increase in the intensity of the corresponding bands with higher B_{12} concentration was detected. The sites a, I, J, L, and M cannot be assigned as the ladders are not sufficiently

resolved in this particular region. Intensity changes at sites 1 (Figure 37 b), 2, 2b, 3, 4 and 8 were plotted *versus* the B_{12} concentration and were fitted to a 1:1 binding isotherm model[97]. The resulting K_D is 150 ± 16 µM, and therefore higher than the value observed for AdoCbl (K_D = 89 ± 6 nM)[87]. A higher K_D value indicates a weaker binding of the B_{12} to the RNA. The K_D of **B_{12}-BHQ-2** is however in the same order of magnitude as other cobalamin derivatives of VitB_{12} (K_D = 314 ± 141 µM)[87].

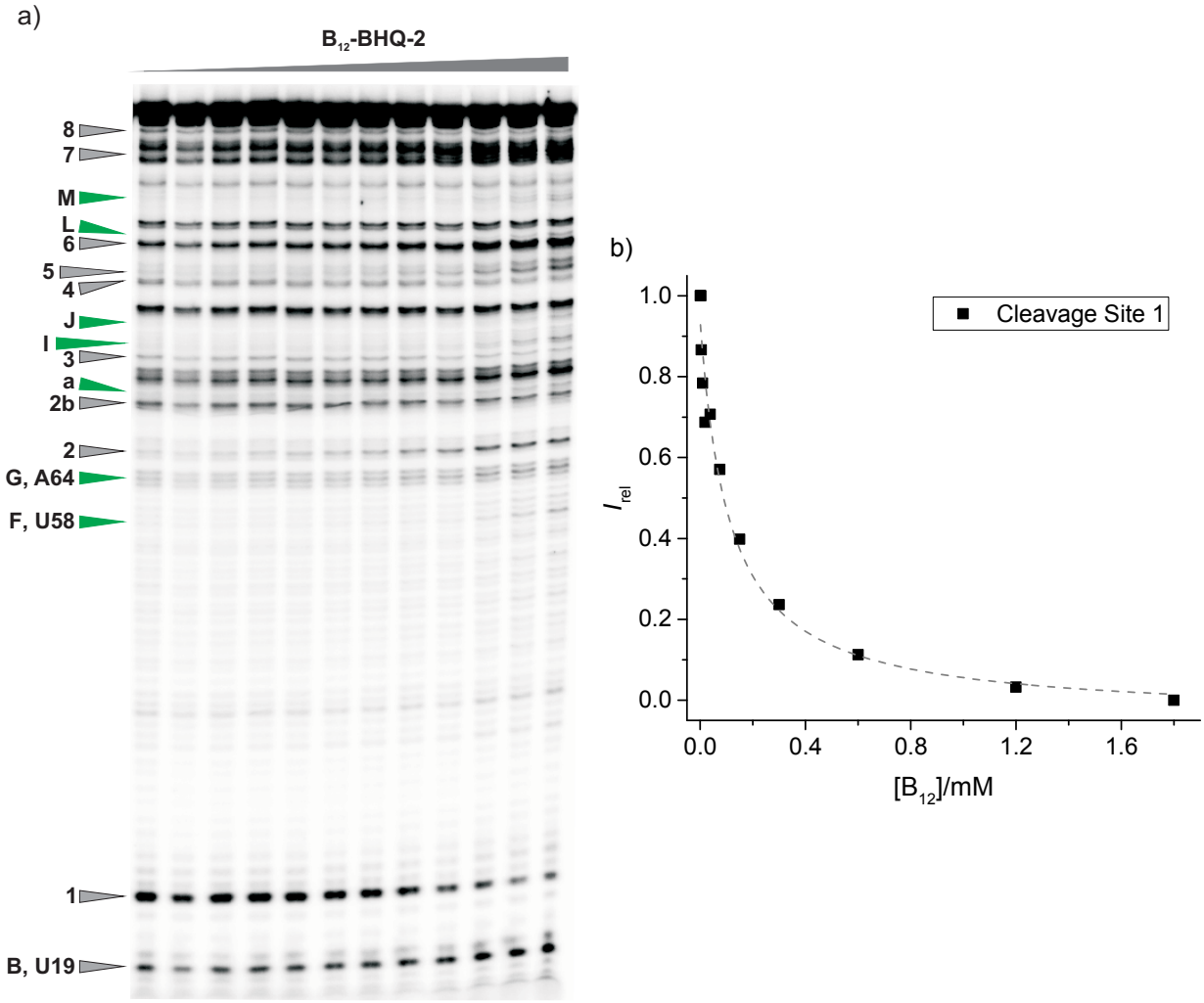


Figure 37: In-line probing experiments of pSG1 with B_{12}-BHQ-2. a) Incubation of the pSG1 riboswitch RNA with increasing concentrations of **B_{12}-BHQ-2**. The characteristic cleavage sites similar to AdoCbl are marked in grey and additional cleavage sites in green. b) Intensity changes at cleavage site 1 of pSG1 in dependence of increasing **B_{12}-BHQ-2** concentrations. Experimental data points are fitted to a 1:1 binding isotherm (dotted line).[97]

The dye-labeled B_{12} derivatives **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** were incubated with the *btuB* riboswitch (*btuB* construct pPC1, Figure 38). Surprisingly, each derivative reveals a different cleavage pattern. The derivatives are fairly similar in their chemical structure with the difference that **B_{12}-Cy3** and **B_{12}-Cy5** are positively charged, and their sulfonated variants **B_{12}-Sulfo-Cy3** and **B_{12}-Sulfo-Cy5** are negatively charged. Cy3 and

Cy5 only differ in two CH_2 groups in the polymethine chain (Figure 36 b). Cleavage sites of each derivative were compared to the ones of AdoCbl. Some cleavage sites are similar to AdoCbl (indicated with grey triangles in Figure 38 a and b) and others are changed (indicated with green triangles in Figure 38 a and b).

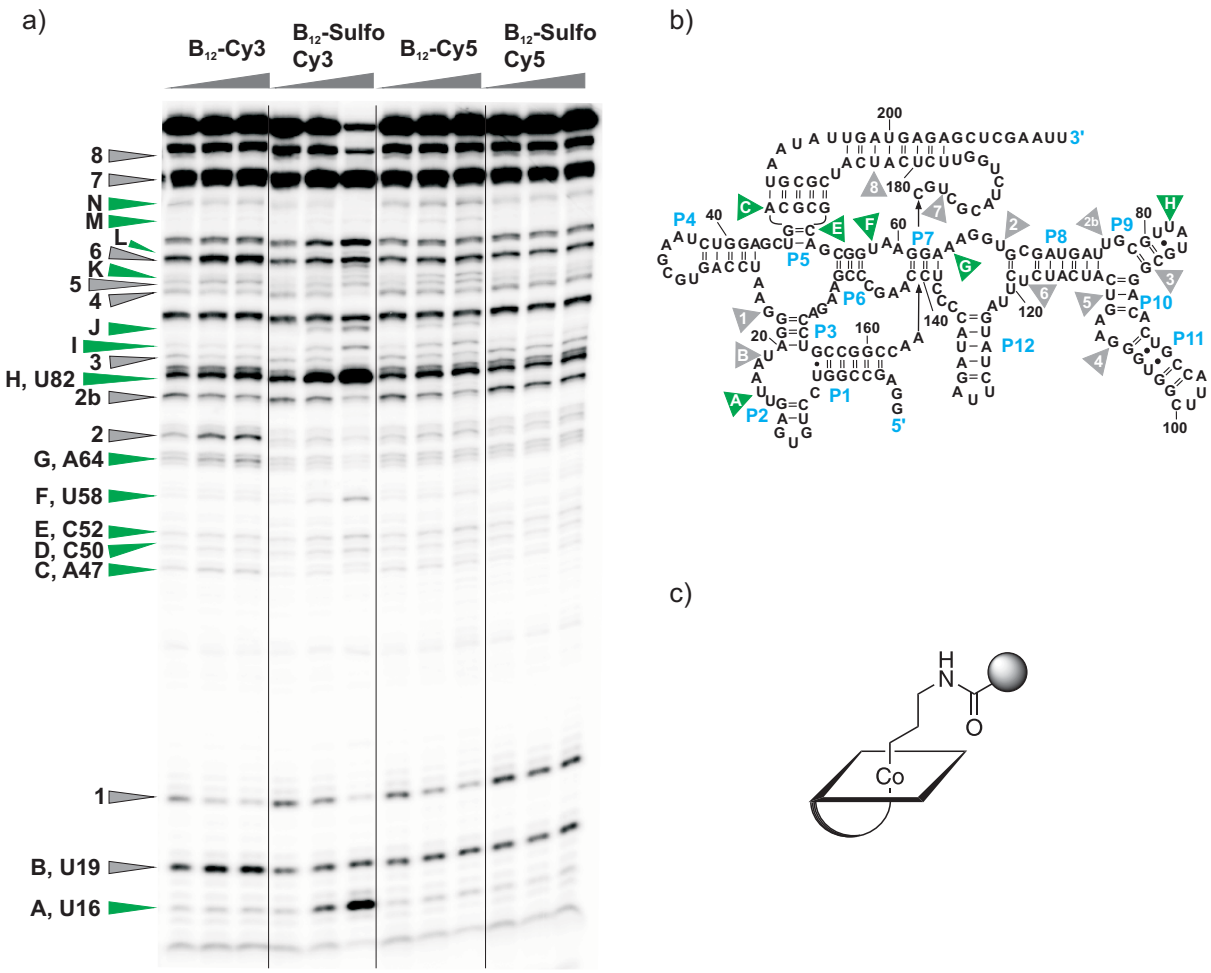


Figure 38: In-line probing experiments of the 216 nt long *btuB* riboswitch (pPC1) with B_{12}-Cy3, B_{12}-SulfoCy3, B_{12}-Cy5 and B_{12}-SulfoCy5. a) Incubation of pPC1 with increasing concentrations of **B_{12}-Cy3**, **B_{12}-SulfoCy3**, **B_{12}-Cy5**, and **B_{12}-SulfoCy5**, respectively. b) Secondary structure of pPC1 including the nine characteristic cleavage sites with AdoCbl (marked in grey) and additional cleavage sites with the B_{12}-dye derivatives (green). The numbering corresponds to the 202 nt long pSG1 structure. c) Schematic structure of the B_{12}-dye derivatives.

Incubation of the *btuB* riboswitch with **B_{12}-Cy3** induces an overall similar cleavage pattern as with AdoCbl (Figure 39 a). However, cleavage sites 5, 6 and 7 are mainly unaffected and reveal no clear trend. Furthermore, three additional cleavage sites C (A47), G (A64), and H (U82) can be observed, which have an increase in intensity with higher B_{12} concentration. Site C is located in P5 (kissing-loop region, see Chapter 2.3.2), G in P7 and H in P9 (see Figure 38 b for the location of the sites). The K_D calculated from the concentration dependent intensity changes at sites 1, 2b, 3, 4 and 8 is 202 ± 42 µM. Compared to **B_{12}-BHQ-2** the binding affinity is in the same order of magnitude.

Incubation of the *btuB* riboswitch with **B_{12}-Cy5** also exhibits mainly similarities to the AdoCbl cleavage pattern (Figure 39 b). Cleavage sites 1, 3, 4 and 8 decrease with higher concentrations of **B_{12}-Cy5**, while 2b and B increase. In contrast, cleavage sites 2, 5, 6 and 7 are unaffected by the B_{12} concentration. This indicates that region P8 is not rearranged upon **B_{12}-Cy5** binding. The additional cleavage site C that was observed for **B_{12}-Cy3** (Figure 39 a) is also detected. Cleavage sites I, J, and M are similar to **B_{12}-BHQ-2** (Figure 37). Additionally, site K, which is situated between sites 5 and 6, reveals an increase in intensity at higher concentrations. Only sites 1 and 8 were fitted to the 1:1 binding model with a resulting K_D of 275 ± 3 µM. The binding constant is therefore similar to **B_{12}-Cy3** and **B_{12}-BHQ-2**.

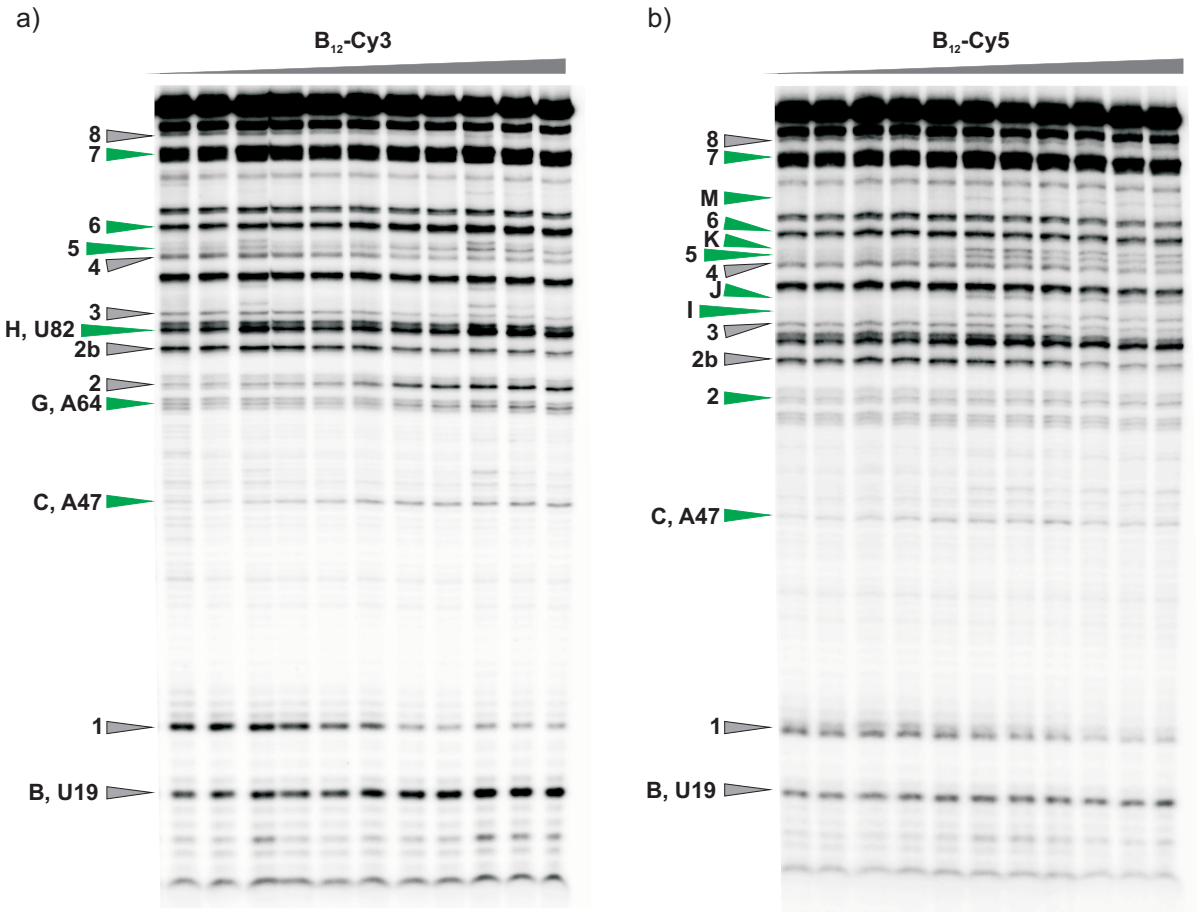


Figure 39: In-line probing experiments of pPC1 with B_{12}-Cy3 and B_{12}-Cy5. Incubation of the pPC1 riboswitch RNA with increasing concentrations of a) **B_{12}-Cy3** and b) **B_{12}-Cy5**. The characteristic cleavage sites similar to AdoCbl are marked in grey and changed or additional cleavage sites in green.

B_{12}-Sulfo-Cy3 induces a different cleavage pattern of the *btuB* riboswitch compared to other derivatives (Figure 40 a). Interestingly, the cleavage pattern is the same as for the *c*-acid VitB_{12} derivative reported by Gallo *et al*. (Figure 40 b).[88]

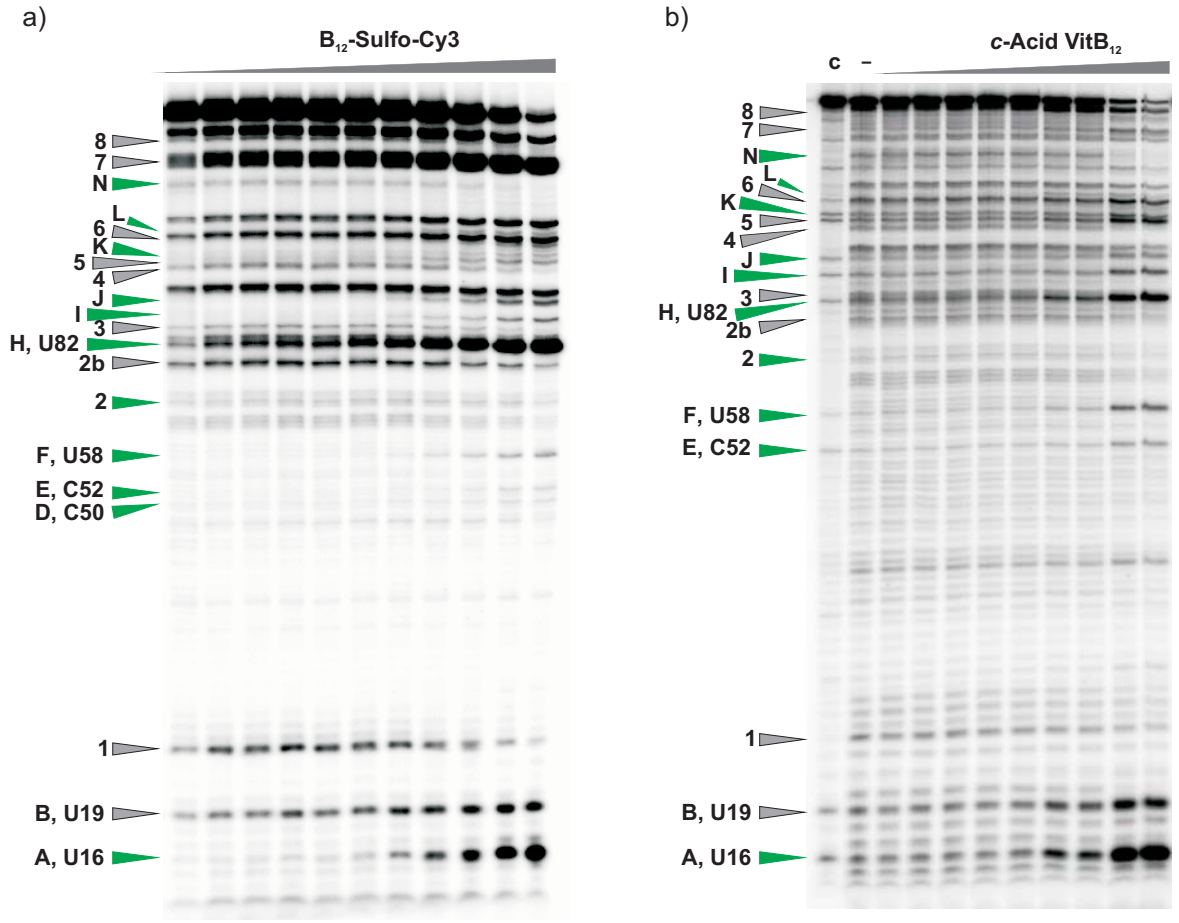


Figure 40: In-line probing experiments of pPC1 with B_{12}-SulfoCy3 and pSG1 with *c*-acid VitB$_{12}$. a) Incubation of pPC1 with increasing concentrations of **B_{12}-Sulfo-Cy3**. b) Incubation of pSG1 with increasing concentrations of *c*-acid VitB$_{12}$ (*adapted from Dr. Sofia Gallo*).[88] The numbering corresponds to the 202 nt long pSG1 structure. The characteristic cleavage sites similar to AdoCbl are marked in grey and changed or additional cleavage sites in green.

Structural similarities between the two B_{12} derivatives exist, which could explain the similar trends (see Figure 41). Both derivatives have a negative charge located at the β-side of the corrin ring; the *c*-acid directly at the corrin ring and the sulfonated Cy3 at the dye in the axial position. An interaction of the negative charges with the riboswitch is therefore likely.

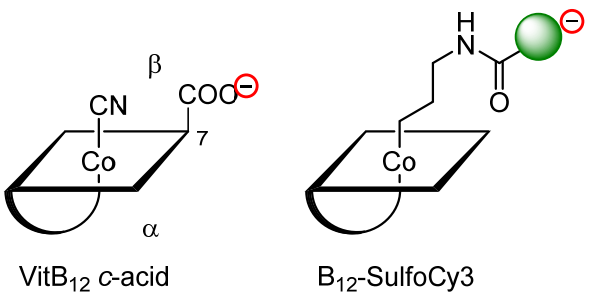


Figure 41: Schematic drawing of the structures of *c*-acid VitB$_{12}$ and B_{12}-Sulfo-Cy3. Both derivatives carry a negative charge at the β-side of the corrin ring (highlighted in a red circle).

Moreover, a site-specific degradation of the RNA is observed with higher **B_{12}-Sulfo-Cy3** concentrations. Therefore most of the characteristic intensity changes are superimposed by this specific degradation pattern. Intensities of cleavage sites 1, 2b, 3, 4, 8 and N are decreasing whereas A, B, D, E, F, H, I, J, 5, K, 6, L, and 7 are increasing with higher concentration of B_{12}. Sites F, I and J were also detected in experiments with **B_{12}-BHQ-2**; I, J, K with **B_{12}-Cy5** and H with **B_{12}-Cy3**. The binding constant of **B_{12}-Sulfo-Cy3** could not be determined as its binding to the riboswitch does not follow a 1:1 binding model. The intensities of the bands rather show a linear decrease or increase.

Incubation of the riboswitch with **B_{12}-Sulfo-Cy5** does not show a significant change in intensity of the cleavage bands, indicating that this derivative does not bind to the RNA at all (Figure 38).

The results of the in-line probing experiments of the *btuB* riboswitch with the B_{12}-dye derivatives are summarized in Table 2 in comparison to naturally occurring B_{12} derivatives.

Table 2: Summarized results of in-line probing experiments of the *btuB* riboswitch with B_{12} derivatives. K_D values of B_{12} derivatives and most notable observations from in-line probing experiments are shown.

B_{12} derivative	**K_D**	**Most notable observations**
AdoCbl	89 ± 6 nM[87]	
CNCbl	314 ± 141 µM[87]	
B_{12}-linker	-	No binding
B_{12}-Cy3	202 ± 42 µM	Additional: C, G, H
B_{12}-Cy5	275 ± 3 µM	Additional: C, I, J, K, M
B_{12}-Sulfo-Cy3		Completely changed pattern
	-	Additional: A, D, E, F, H, I, J, K, L, N
		Site-specific degradation at high [B_{12}]
B_{12}-Sulfo-Cy5	-	No binding
B_{12}-BHQ-2	150 ± 16 µM	Additional: a, F, G, I, J, L, M

2.6 Discussion and Conclusion

The driving force for the above described studies was the investigation of a new labeling strategy for the elucidation of the structural rearrangements of the *btuB* riboswitch upon binding of the B_{12} metabolite by FRET. To this end, dyes of the common FRET couples Cy3/Cy5 and Cy3/BHQ-2 were chemically attached to the fairly large metabolite itself. New B_{12}-dye derivatives were successfully synthesized and characterized.

A first step was the synthesis of the fluorescent cyanine dyes Cy3 and Cy5 followed by their activation to an NHS ester. The one-pot synthesis of Cy3 resulted in a low yield compared to the multi-step approach known from literature due to the formation of various side products.[93] An alternative synthetic approach is therefore recommended, *e.g.* a one-pot synthesis postulated by Korshun and coworkers in which Cy3 can be obtained in a higher yield.[27]

To find the optimal labeling scheme which induces a correct folding of the riboswitch, derivatives of Cy3 and Cy5 differing in their substituents and charges were used. To determine the optimum measuring conditions for FRET additionally the FRET pair was varied. The donor dye Cy3 is placed at the B_{12} and either one of the acceptor dyes Cy5/BHQ-2 at the RNA, or vice versa. The attachment of the dye was chosen to be in axial position of the cobalamin since this position of the cobalamin should only influence the binding affinity of the B_{12} to the riboswitch, whereas modifications at the corrin ring can lead to a different folding of the RNA (see Chapter 2.3.4).[87] The dyes were coupled via a covalently bound linker to the cobalt center of the B_{12} yielding the new derivatives **B_{12}-Cy3**, **B_{12}-Cy5**, **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** and **B_{12}-BHQ-2**.

Stability measurements, exemplarily undertaken with **B_{12}-Cy3**, showed a partial decomposition to B_{12} and Cy3 due to the Co-C bond cleavage upon excitation at 532 nm for 10 min. This is due to the low dissociation energy of ~37 kcal/mol of the Co-C bond that is within the range of visible photons.[53] The photolysis could therefore be a disadvantage in FRET experiments using laser excitation of Cy3 at 532 nm. However, by keeping the measuring time short and the excitation intensity low enough this effect is negligible.

Another issue of B_{12} derivatives is the fluorescence quenching ability of cobalamin of the excited electronic state of the attached fluorophore. The electronic orbital of cobalamin overlaps with the frontier orbital of the excited state of the fluorophore.[94] Fluorescence quantum yields of B_{12}-flurophore derivatives are therefore lower than of the free dyes. To overcome the quenching, Grissom and coworkers reduced the intramolecular dynamic movement by incorporation of a rigid linker, which is attached at the ribose moiety of the cobalamin, and thus further away from the corrin ring.[98] However, this approach is not suitable with the *btuB* riboswitch, as it is known that modifications at this position would lead

to an unnatural folding of the RNA, which can lead to a loss of function. On the other hand, one could take advantage of the quenching behavior of the B_{12} and use it for FRET studies acting as acceptor. This approach is similar to the use of dark quenchers (Chapter 1.2.2).[99,100]

As the rotational freedom of the dyes is an important factor in terms of FRET (see Chapter 1.2), steady-state fluorescence anisotropy measurements were conducted. Similar values for free Cy3 and Cy3 attached to B_{12} were observed. Comparison of free Cy5/Sulfo-Cy5 to **B_{12}-Cy5**/**B_{12}-Sulfo-Cy5** shows an increase of the anisotropy value r by 0.1. A high value is usually associated with a rotationally restricted environment. Attachment to a macromolecule typically increases the r value of the dye.[101] However, the origin of the different behavior of Cy3 and Cy5 remains unclear. A possible reason is that steady-state fluorescence anisotropy in some cases can give misleading results, as it has been shown by Levitus and coworker.[21] In a next step time-resolved measurements could give more details about the rotational freedom of the dye upon attachment to B_{12}.

Furthermore, in-line probing experiments were performed in order to clarify the structural conformation of the *btuB* riboswitch in presence of the modified B_{12} derivatives **B_{12}-linker**, **B_{12}-Cy3**, **B_{12}-Cy5**, **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** and **B_{12}-BHQ-2**. The observed cleavage sites were compared to the ones of the natural metabolite AdoCbl. Influence of the dyes by themselves on the RNA was ruled out by control experiments in which the RNA was incubated in presence of the free dyes. To compare the influence of the replacement of the upper ligand (adenosyl moiety) by the amine linker, the RNA was incubated with **B_{12}-linker**. No significant changes in the intensities of the cleavage bands were observed, suggesting no binding of this derivative under the applied conditions. It is known from the literature that replacing the bulky upper ligand by a smaller residues leads to a reduced binding affinity to the RNA.[87] Additionally, under the applied conditions a positive charge is located at the primary amine of the linker which might affect the binding. Altogether, the binding affinity of **B_{12}-linker** is too low and no specific binding to the riboswitch occurs. The binding affinities of **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-BHQ-2** are in the order of magnitude of $VitB_{12}$ (~ 200 µM) and therefore 1000 times lower than the one of AdoCbl. The consequence for FRET experiments is the need of higher B_{12} concentrations, which increases background fluorescence. To overcome this drawback zero-mode waveguides or vesicle encapsulation could be used to lower the fluorescent species concentration.[102,103] Furthermore, possible changes in the RNA folding behavior as a consequence of the structural changes to the natural metabolite have to be considered. Although the dyes have a similar chemical structure, each B_{12}-dye conjugate induces a different folding behavior of the riboswitch. **B_{12}-Cy3** induces an additional cleavage site C which might be an indicator that the kissing loop interaction is affected. From the crystal structure of another B_{12}-binding riboswitch is known that these

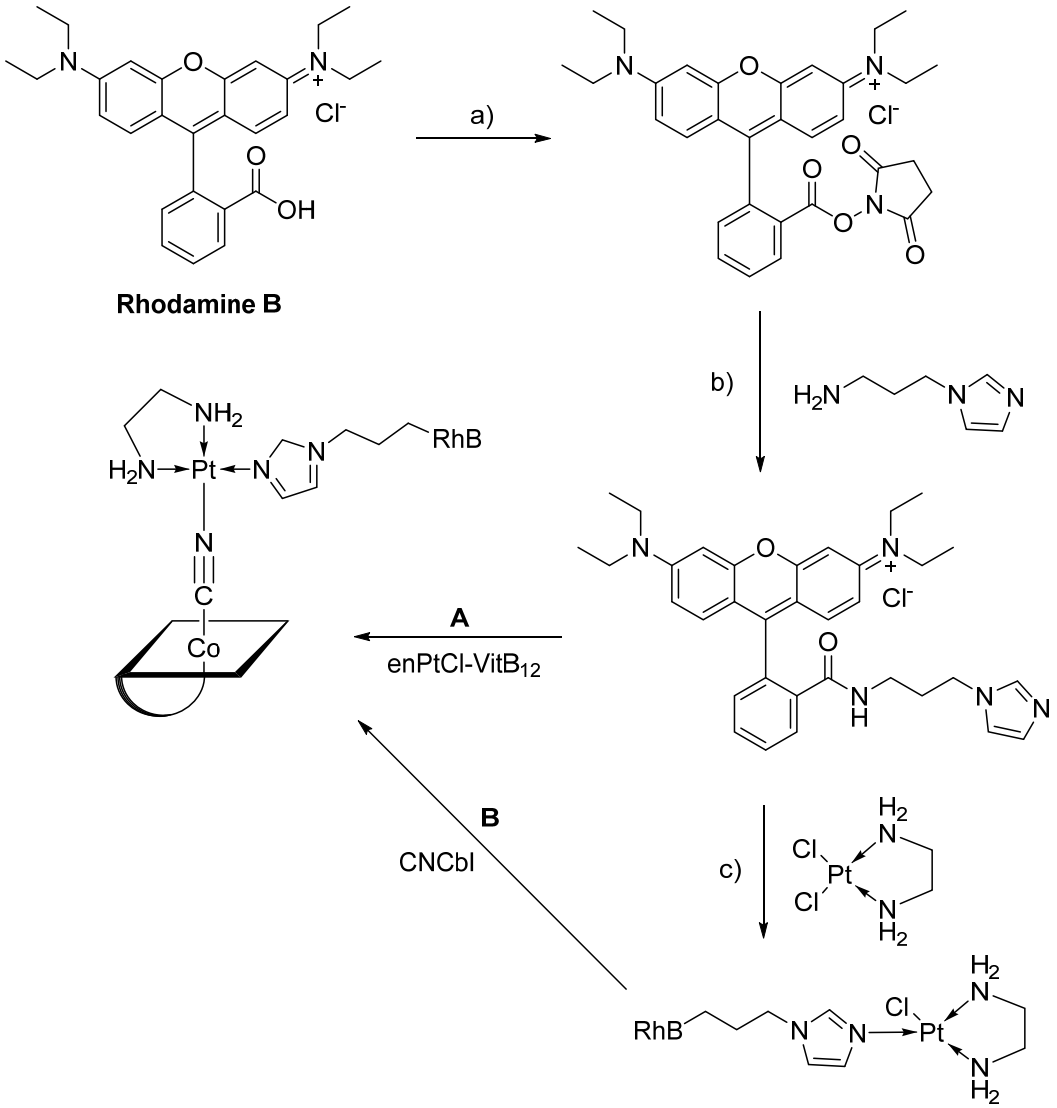
motif encapsulates the cobalamin in the binding pocket.[83] Consequently, Cy3 might sterically prevent the complete formation of this interaction. Nevertheless, **B_{12}-Cy3** shows a rather similar behavior to AdoCbl. In contrast, the sulfonated variant **B_{12}-Sulfo-Cy3** exhibits a completely changed cleavage pattern. The changes are analogous to the ones previously observed for *c*-acid VitB_{12}.[88] In both cases site-specific degradation of the RNA occurs at high B_{12} concentrations. The *c*-acid carries a negative charge at the β-side of the corrin ring. The negative charge of the sulfonated Cy3 in **B_{12}-Sulfo-Cy3** is also located at the dye in the axial position of B_{12}. This clearly indicates an influence of charged ligands on the riboswitch folding due to electrostatic effects. **B_{12}-Cy5** and **B_{12}-BHQ-2** induce some additional cleavage sites compared to AdoCbl, but show an overall similar cleavage pattern. Unexpectedly, incubation of the riboswitch with **B_{12}-Sulfo-Cy5** does not induce significant changes in the intensities of the cleavage bands, suggesting that this derivative does not bind to the RNA. Absence of binding due to steric hindrance would be expected to also affect the unsulfonated derivative **B_{12}-Cy5**. If the negative charge would lead to a charge repulsion and, therefore, prevent the ligand from binding to the RNA, the same would be true for the sulfonated Cy3 derivative **B_{12}-Sulfo-Cy3**. It seems that the combination of steric and electrostatic interactions causes that **B_{12}-Sulfo-Cy5** is not recognized by the riboswitch. Consequently, charge, size and position of modifications on the B_{12} are crucial for the folding of the riboswitch, showing once more the ligand specificity of the riboswitch.

In conclusion, the **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-BHQ-2** are the most promising candidates for FRET experiments. The attachment of bulky residues like the fluorescent dyes is slightly affecting the RNA folding, but in-line probing experiments indicate that these changes are fairly small. The binding affinity of these derivatives is reduced compared to the natural ligand AdoCbl, but still in the range of VitB_{12}. **B_{12}-Sulfo-Cy3**, and **B_{12}-SulfoCy5** are no suitable probes to study the folding behavior of the riboswitch with FRET. Nevertheless, both derivatives contributed to new insights helping to refine the picture of the riboswitch folding and interaction with the B_{12} metabolite. Furthermore, the new derivatives can be used for other fluorescence applications such as imaging of cobalamin receptors or pathways of B_{12}.

2.7 Outlook

With the presented labeling strategy also other dyes can be attached to the B_{12}. Like this fluorescence quantum yields could be further improved. A smaller and uncharged dye could even less affect the folding of the riboswitch.

To address the problem of light-stability of the B_{12}-dye compounds light stable derivatives which are not based on linkage by a Co-C bond can be designed. The Co-CN bond in CNCbl ($VitB_{12}$) is known to be stable upon light irradiation. The cyanide ligand can then be further derivatized, *e.g.* by coordination of $enPtCl_2$ to give $[enPtCl\text{-}VitB_{12}]^+$ to which additional groups can be coordinated by replacement of the chloride ligand (Scheme 5).[104]



Scheme 5: Synthetic route of a light-stable B_{12}-RhB derivative. a) DCC, NHS, acetonitrile, b) carbonate buffer at pH 8.3, c) $AgNO_3$, DMF.

The new route can be developed by using a cheap fluorophore such as rhodamine B and can later on be applied with other dyes. A synthetic pathway is proposed in Scheme 5.

Rhodamine B (RhB) is extended with aminopropylimidazole as linker group for a specific coordination to platinum using *N*-hydroxysuccinimide (NHS) ester chemistry.[105] In route A the modified RhB is coordinated to enPtCl-VitB_{12}, in route B to enPt(II)Cl_2 followed by coordination to vitamin B_{12}. Reaction steps a) through c) have already been realized in preliminary experiments. In both pathways a last step is expected to lead to the final product.

For further investigations of the *btuB* riboswitch of *E.coli* with the B_{12}-dye derivatives (especially **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-BHQ-2**) by FRET, the RNA needs to be fluorescently labeled as well with either Cy3 or Cy5 to complete the FRET pair. A possible strategy to address this issue is given in Chapter 3.

3 Labeling of RNA

3.1 Introduction

Fluorophore labeling of large biomolecules such as proteins, RNAs or DNAs is the basis for a vast variety of applications especially in the field of fluorescence spectroscopy. Fluorescent labels can be used for imaging of pathways, cellular processes, and folding mechanism by ensemble and single-molecule fluorescence methods. One of these techniques is Förster Resonance Energy Transfer, a powerful tool to study the conformational dynamics of biomolecules, especially on the single molecule level (Chapter 1.2).[35,36] The site-specific internal labeling of large (> 200 nucleotides, nts) functional RNA molecules is challenging and often not practicable. Indirect labeling can be achieved by hybridization of fluorescently labeled DNA oligonucleotides to the RNA.[106] The introduction of modifications such as long loops within the RNA structure are necessary for an efficient binding of the label. Those modifications can disturb the structure and thus the functionality of the RNA.
In the following an RNA system is introduced, a group II intron ribozyme, which serves as model system to investigate a new strategy for the fluorescence labeling of large RNAs. The folding behavior of this system is known with the commonly used labeling strategy and can therefore be used to evaluate the potential of the new strategy.

3.2 Ribozymes

As described previously, one function of RNA is the regulation of the gene expression, *e.g.* in the form of riboswitches (Chapter 2.3). Further functional RNAs are ribonucleic acid enzymes, so-called ribozymes, which are catalytically active.[107] For their discovery in 1982 S. Altman and T. R. Cech were awarded with the Nobel Prize in Chemistry in 1989.[108,109] Since then the research interest in this field was increasing drastically.
Several small ribozymes (40-160 nts) exist that are well studied and act as ideal model systems to study their structure function relationship like the hammerhead,[110] hairpin,[111] and the hepatitis delta virus (HDV) ribozyme[112] (Figure 42).[113] Large ribozymes (400-3000 nts) such as the group I and II introns are more difficult to study due to their complex structure. Introns and exons are nucleotide sequences contained in genes. Intron sequences are the parts, which do not encode the protein sequence and interrupt the coding exons. They are removed by RNA splicing prior to translation. Afterwards only the joined exons are remaining to generate the protein. Introns can be classified in four categories: introns in nuclear protein-coding genes removed by the spliceosome, introns in tRNA genes removed by proteins,[114] self-splicing group I introns, and self-splicing group II introns.

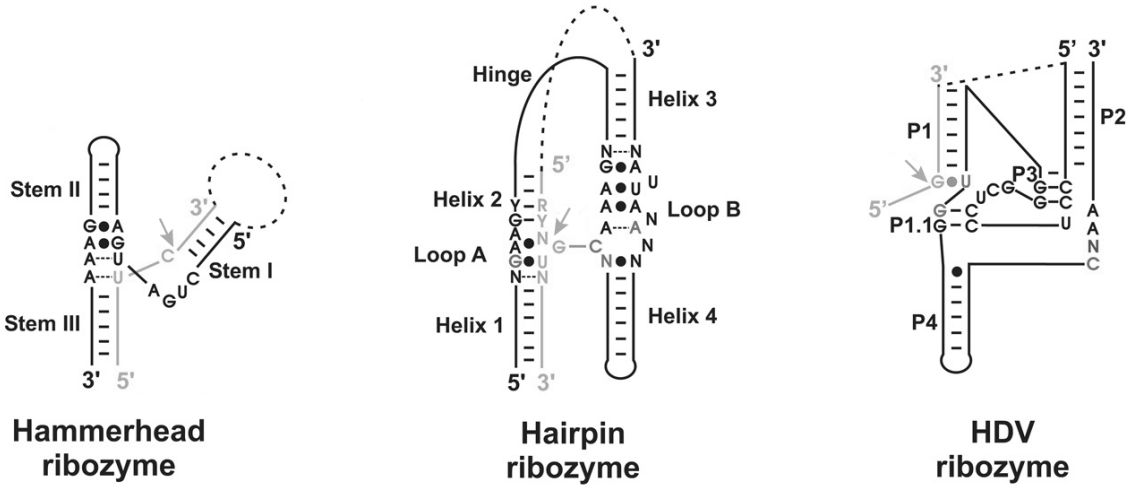


Figure 42: Schematic representation of the structures of different small ribozymes. Schematic structures of the hammerhead, hairpin and HDV ribozyme. Substrates are depicted in grey, cleavage sites are indicated by arrows. *Figure adapted from Ref. [115]*.

3.2.1 Classification and Function

Ribozymes are a good example for the structure-function relationship. These catalytic RNA molecules adopt a certain geometry to support their activity. This characteristic is typically assisted by metal ions. Ribozymes can catalyze a diversity of reaction types such as phosphoryl transfer or peptidyl transferase as for naturally occurring ribozymes. *In vitro*-selected species can for example catalyze alkylation, isomerization, metal transfer, and Diels-Alder cycloaddition.[115,116]

Most of the ribozymes perform phosphoryl transfer with transesterification reactions, which are therefore in the following described more in detail. Transesterification is the exchange of an organic group of an ester with the organic group of an alcohol. Depending on the respective reaction mechanism the ribozymes have been categorized into different classes. The reaction is initiated by a nucleophilic attack on a phosphate by:

- the adjacent 2' hydroxyl *(nucleolytic ribozymes)*.
- the 3' hydroxyl of an exogenous guanosine *(group I introns*).
- a remote nucleotide *(group II introns).*

Moreover, ribozymes can be 'self-cleaving' or 'self-splicing'. Self-splicing ribozymes such as introns can cut themselves out of larger RNAs and ligate the cut ends.[107] Others, such as RNase P, cut off a sequence or precursor on tRNA.[109,117,118]

The cleavage reactions can be achieved in general by two different mechanisms with different intermediates. Small ribozymes are self-cleaving. Their catalytic function is based on a site-specific internal reaction mechanism with a cleavage of the phosphodiester bond by an internal reaction from the 2'-hydroxyl of the sugar. Products formed in this reaction are a

2',3'-cyclic phosphate as well as a 5'-hydroxyl terminus (Figure 43 a). The reverse reaction is a ligation involving two proton transfers. Although all small ribozymes catalyze the same chemical reaction of specific backbone cleavage, they differ mechanistically and structurally.[113] Large ribozymes catalyze by a mechanism involving an external nucleophile, typically a hydroxyl group, to cleave the phosphodiester in the backbone (Figure 43 b). RNase P for example catalyzes the hydrolysis reaction with water as nucleophile.[115]

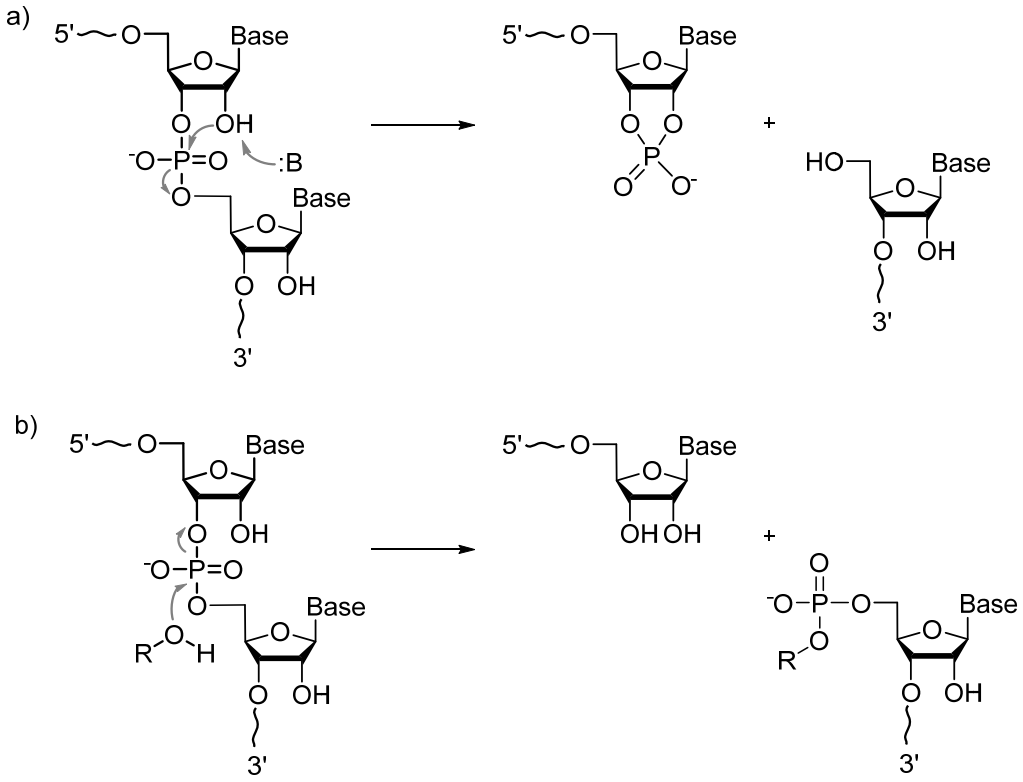


Figure 43: RNA cleavage mechanisms. a) Cleavage of phosphodiester linkage initiated by a nucleophilic attack of a base (B) at the 2'-hydroxyl of the sugar. The cleavage products are a 2',3'-cyclic phosphate and a 5'-hydroxyl terminus. b) Hydrolysis (R = H for RNase P) or transesterification (R = organic residue, guanosine for group I introns, internal adenosine for group II introns).[115]

Different functional elements can be involved in the ribozyme catalysis. One is the RNA structure which is responsible for the position and orientation of the corresponding substrate. Metal ions are involved, besides their structural role, as Lewis acids or source of water molecules for acid-base catalysis. Furthermore, nucleobases act as acids, bases or via electrostatic effects.[116,119]

Splicing

As mentioned above introns are non-coding sequences which are removed to give a functional RNA sequence by different mechanisms. The 'self-splicing' group I and II introns vary in the splicing mechanism by the attacking nucleophile. In group I introns the nucleophile is the 3'-OH of a guanosine residue whereas in case of group II introns of an internal adenosine.

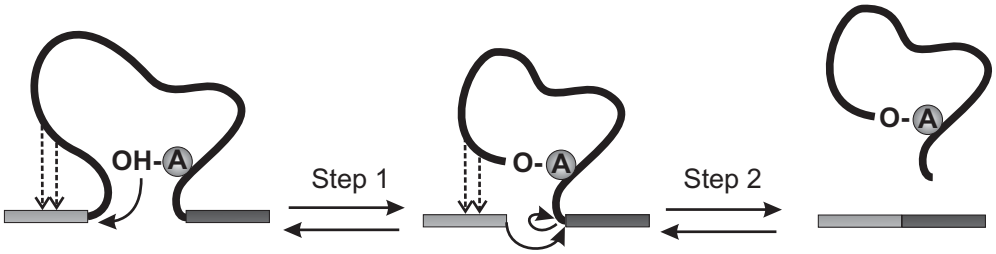


Figure 44: Splicing reaction of group II intron ribozymes. Step 1 is a nucleophilic attack by the 2'-OH of an internal adenosine. In step 2 the newly generated 3'-OH of the 5'-splice site attacks the phosphate bond at the 3'-splice site. The intron (solid black line) is released as lariat and the exons (grey bars) are ligated.[120]

The splicing reaction of group II intron ribozymes includes two transesterification steps. Step 1 is a nucleophilic attack by the 2'-OH of an internal adenosine. In step 2 the newly generated 3'-OH of the 5'-splice site attacks the phosphate bond at the 3'-splice site. The intron is released as lariat and the exons are ligated (Figure 44). Tertiary interactions of the exon and intron binding sites (EBS-IBS) are crucial for the reaction. Typically the splicing reaction is reversible. However, besides this branching pathway alternative pathways exist based on hydrolysis which are irreversible. In this case a linear intron is formed rather than a cyclic lariat.[121–124]

3.2.2 Group II Intron Ribozymes

Group II introns are among the largest known catalytic RNA molecules (500 to > 1000 nts) and are found in bacteria and in organelles of various eukaryotes, especially fungi and plants, but not in animal genomes.[125] Depending on their secondary structure group II introns can be further divided into subclasses IIA, IIB, and IIC.[126,127] They have in general a conserved global secondary structure typically organized in six domains around a central wheel structure with several long-range tertiary interactions between the different domains (Figure 45).[124,127] Domain 1 (D1) is the largest domain and is essential for the recognition of the exonic substrate.[128,129] It forms a folding scaffold and provides key active-site elements. Together with domain 5 (D5), the catalytic center, D1 is crucial for the catalytic activity.[130,131] Domains 2 and 3 enhance the catalytic activity but are not essential.[132] D4 is the domain that varies most and can contain an open reading frame for translation of a multifunctional

protein involved in splicing.[133] D6 includes the adenosine residue branching point for the splicing.[121]

Relevant tertiary interactions such as α-α', ζ-ζ', κ-κ', λ-λ' between the different domains help to compose the intron active site.[127,134–137] The exon binding site 1 (EBS1) is positioned within the catalytic heart whereas EBS2 is completely exposed.[129,138] The key interactions of the group II intron architecture have mainly been characterized in biochemical studies. Furthermore, single domains, tertiary interactions and metal-ion binding sites were elucidated by NMR.[139–141] It has only been recently that a group II intron was successfully characterized by X-ray crystallography.[142] However, crystallization of such large and complex RNA structures remains to be challenging and time-consuming. Exemplarily, a three-dimensional crystal structure of the *Oceanobacillus iheyensis* intron is shown in Figure 46.[142]

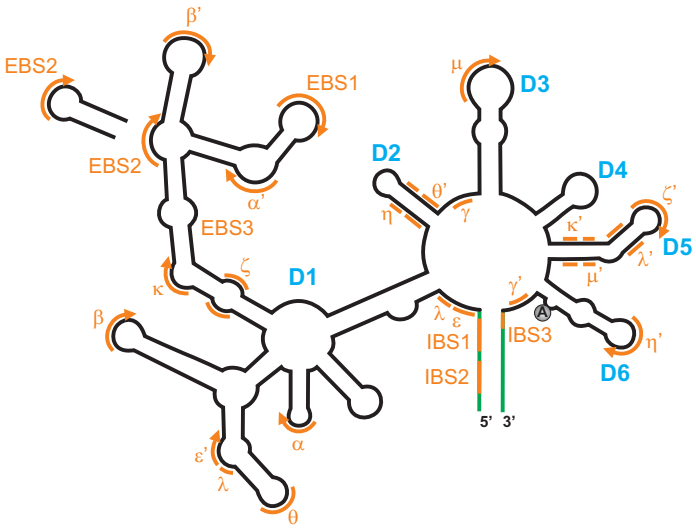


Figure 45: Schematic structure of group II introns with characteristic tertiary interactions. Group II introns (black) are organized in six domains (blue) and are located in the exonic sequence (green). Tertiary interactions (orange) are highly conserved.[124,125]

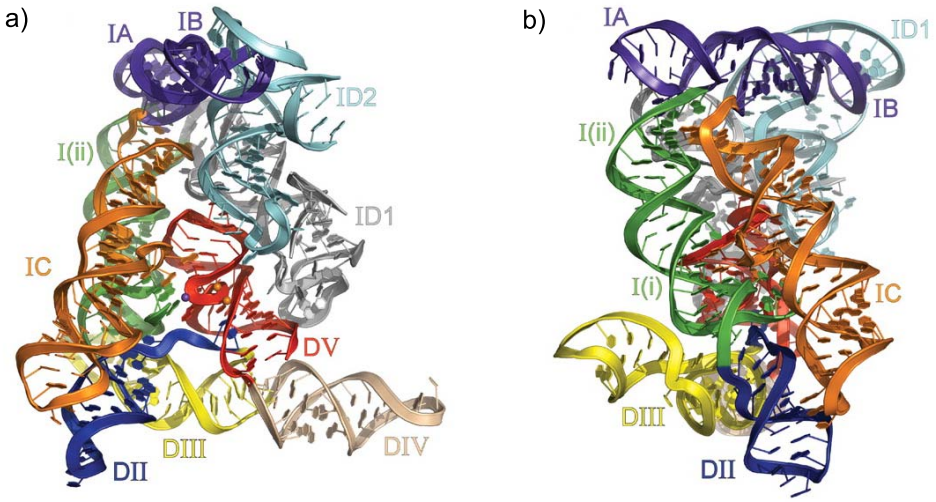


Figure 46: Crystal structure of the *O. iheyensis* intron. a) Overall structure of the intron in a ribbon representation. b) 90° rotation of the image shown in a). (*Figure from Ref. [142]. Reprinted with permission from AAAS*).

Most of the tertiary interactions discussed above can indeed be observed in the structure of the *Oceanobacillus iheyensis* intron. This intron belongs to the group IIC introns which are highly reactive and smaller than the IIA and IIB introns. Furthermore, group IIC introns are supposed to splice via a hydrolytic pathway forming a linear intron instead of a cyclic lariat product (see Chapter 3.2.1).[142]

3.2.3 Group II Intron *Sc*.ai5γ

The mitochondrial group II intron ai5γ from *Saccharomyces cerevisiae* (*Sc*) is a well-studied system. This 900 nts large ribozyme self-splices by changing its structure.[124,125,143–145] The intron interrupts the cytochrome oxidase subunit I gene.[129,134] Many discoveries relevant for the tertiary interactions of group II introns in general are based on ai5γ. *Sc*.ai5γ is therefore a very interesting system to study the folding behavior of large RNA molecules.[42,143]

A common approach to study large and complex systems is the use of smaller constructs as model systems. Results obtained with these model constructs can then be transferred back to the large systems. D135 is a deletion construct of ai5γ containing all components contributing to the active site. It folds into a uniform catalytically active compact structure and is able to perform the first step of the splicing pathway. D135 cleaves a short oligonucleotide complementary to the exon binding sequences EBS1 and 2. The reaction follows thereby a hydrolytic rather than a branching reaction (Chapter 3.2.1). Pyle and coworkers determined the fraction of active ribozyme molecules by active-site titration experiments under multiple turnover conditions where the product release is rate limiting with an excess of substrate. The conditions for highest catalytic activity were determined. It was found that 83 ± 5% of the D135 population is active at 42 °C. The best conditions for ensuring complete folding of ribozyme *in vitro* were found to be 100 mM $MgCl_2$, 500 mM KCl at pH 7.0. For this reaction the burst rate (beginning of reaction before steady state is reached) of 0.48 ±0.1 min^{-1} is equivalent to the single turnover rate with the burst population being the most active state of ribozyme.[138,145]

To elucidate the structural dynamics of the group II intron folding by smFRET Steiner *et al.* designed a new construct D135-L14 (Figure 47) which is based on D135. Two artificial loops L1 and L4 were additionally introduced into domains 1 and 4. These modifications allow the specific hybridization of two DNA oligonucleotides carrying the FRET dye pair Cy3 and Cy5 (see Chapter 3.3). Furthermore, the 3'-end was extended for a hybridization of DNA-biotin to immobilize the construct to a surface via biotin-streptavidin linkage.

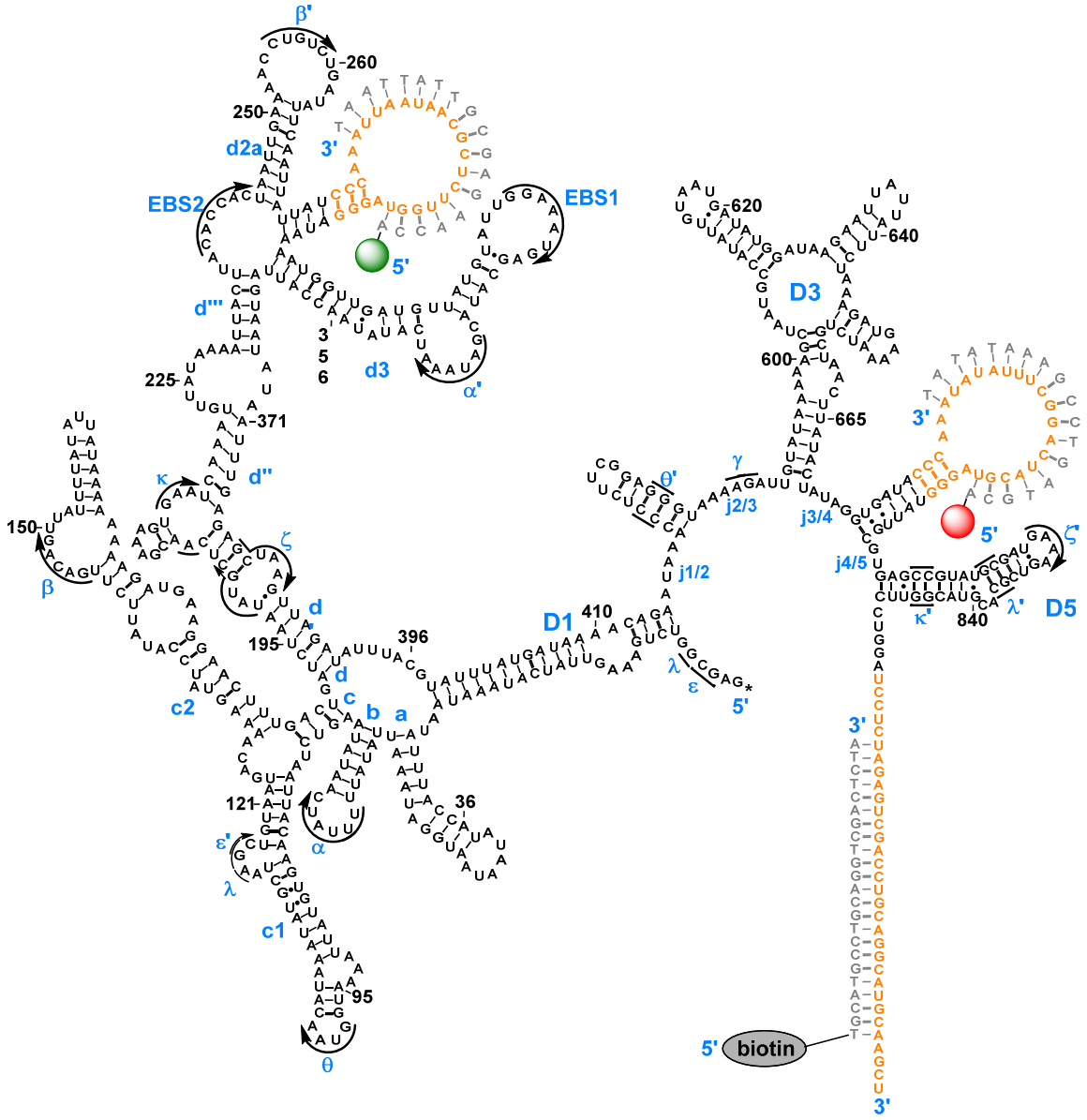


Figure 47: Secondary structure of D135-L14 with hybridized DNA-labels. DNA oligonucleotides (grey) carrying Cy3 (green), Cy5 (red) or biotin hybridized to D135-L14 RNA. The numbering corresponds to the wild-type *Sc*.ai5γ sequence and the tertiary interactions are indicated in Greek letters. Structural modification to D135 are marked in orange.[42]

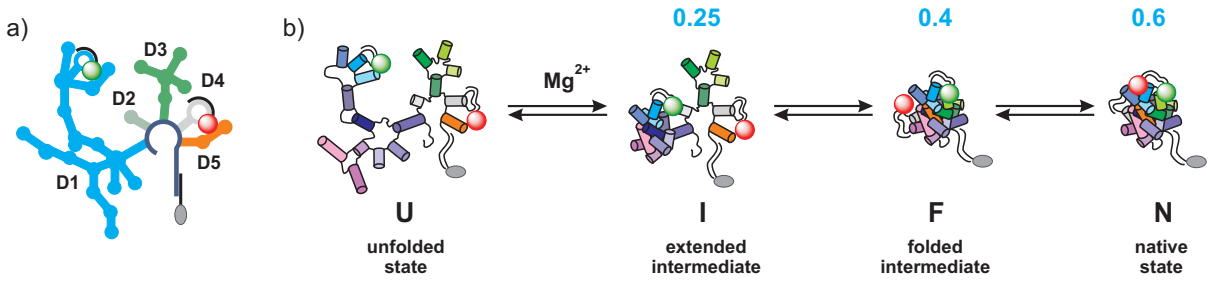


Figure 48: Schematic folding pathway of a group II intron. a) Schematic structure of D135-L14 with Cy3 (green sphere), Cy5 (red sphere) and biotin (grey sphere) labels. b) Folding pathway of D135-L14. *Figure adapted from Ref. [42].*

Single-molecule studies *in vitro* revealed that the group II intron folds sequentially in three steps from the unfolded state U via intermediates I and F to a compact native state N (Figure 48). Mg^{2+} has been found to be crucial for the folding. The native state N is stabilized upon substrate binding and has been proposed to be the catalytically active species.[42]

3.3 Labeling of Large RNAs

An efficient fluorescence labeling of the molecules of interest is the base for various fluorescence techniques, *e.g.* FRET (Chapter 1.2). Thus, the fluorophore labeling sites have to be chosen carefully for the investigation of molecule dynamics, as they are sensitive to relevant folding transitions. In particular, the labeling should not significantly perturb the folding or kinetics of the molecules under investigation.

The site-specific internal labeling of large (> 100 nts) functional RNA molecules is challenging. Modifications within large RNAs are mainly post-synthetically, in contrast to shorter RNAs up to 100 nts in length where modified bases or reactive functionalities can be incorporated during solid phase synthesis.[146,147] A direct internal labeling is often not specific. To address the problem of specificity, direct post-transcriptional labeling methods involve the use of guides, *e.g*. complementary oligonucleotides, to introduce functionalities within the RNA. Another approach is the use of enzymes such as deoxyribozymes for the attachment of fluorescent reporter groups.[148,149] In general, those methods are mostly limited by an extensive preparation of the guide, further purification steps of the labeled molecule and consequently low labeling yields. Alternatively, large RNAs can be synthesized from separate short RNA sequences by ligation. This allows functionalizing small RNAs at the 3' or 5' terminus, respectively, or site-specifically label them following the established techniques for small RNAs (*vide supra*). However, it was observed that this approach can lead to a loss of function.[147,150]

Pan and coworkers developed a method for indirect labeling by modification of the RNA with single stranded loops or extensions, to which complementary DNA oligonucleotides carrying the label can be hybridized.[106] To ensure the formation of a stable RNA-DNA duplex and a

sufficient labeling efficiency, an overlap of at least 18 nts is necessary implying large RNA modifications, which can influence the folding behavior. Alternatively, stronger binding non-natural nucleic acids such as peptide nucleic acids (PNAs) could be used to overcome these drawbacks. PNA is already known from other applications to have several advantages compared to DNA. In the following will be evaluated if PNAs are suitable labeling probes to replace DNA oligonucleotides for hybridization to RNA.

3.4 A Non-Natural Nucleic Acid Analog: Peptide Nucleic Acid

In 1991 Nielsen, Egholm, and coworkers first invented/discovered the non-natural nucleic acid derivative peptide nucleic acid (PNA).[151] It consists of an uncharged pseudo-peptide backbone whereas DNA and RNA have the negatively charged phosphate groups. PNA binds therefore stronger to RNA than either DNA or RNA do, forming very stable Watson-Crick paired complementary duplex structures. PNA serves as a structural mimic of DNA and RNA and was originally designed as a ligand for the recognition of double stranded DNA.[152]

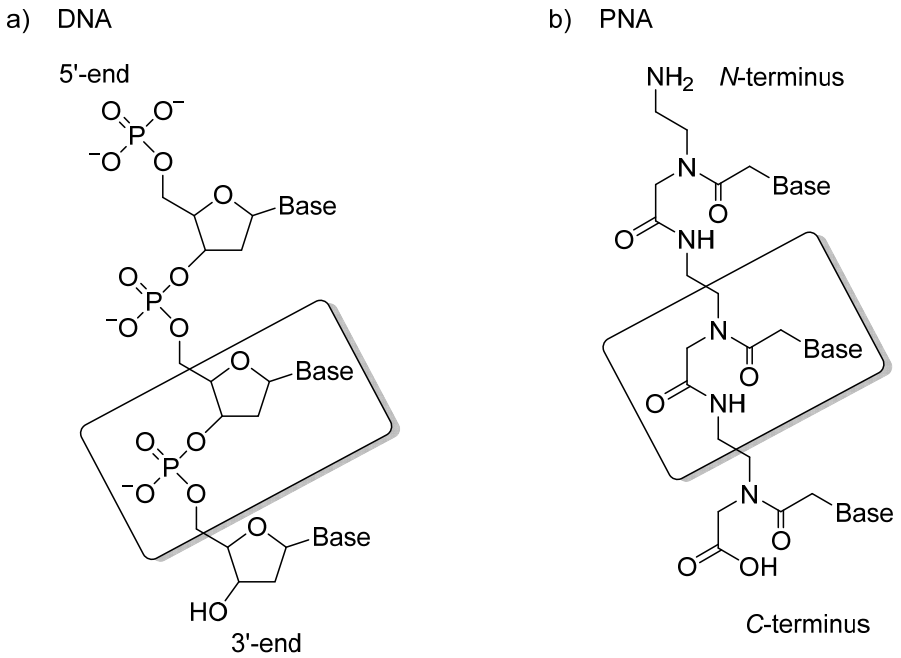


Figure 49: Comparison of chemical structures of DNA and PNA strands. a) DNA and b) PNA with characteristic units boxed in. DNA consists of a negatively charged sugar phosphate backbone. PNA has an uncharged pseudo-peptide backbone.

The structure of PNA consists of repeating *N*-(2-aminoethyl)-glycine units linked by amide bonds. The bases are attached to the backbone by methylene carbonyl linkage. Unlike DNA or RNA no pentose sugar moieties or phosphate groups are present, which results in an acyclic, achiral and neutral backbone (Figure 49). The consequences of this neutral backbone lead to the important feature of PNA, its strong binding between complementary

PNA-DNA and PNA-RNA strands at low to medium ionic strength. The binding strength is stronger than between DNA-DNA or DNA-RNA due to the lack of charge repulsion.

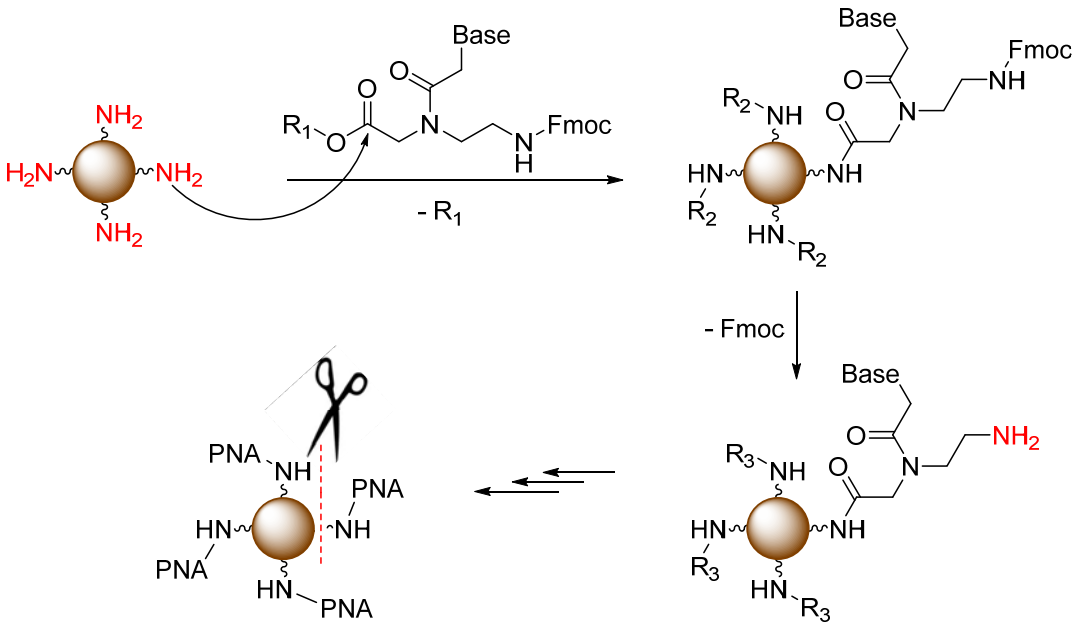
Furthermore, PNA has a much higher sequence specificity as mismatches in PNA-DNA or PNA-RNA duplexes are more destabilizing than in DNA-DNA or RNA-RNA.[153] Watson Crick base pair rules are strictly observed in the hybrids.[154] Like this, PNA can bind to complementary nucleic acids in an antiparallel orientation (*C*-terminus and 5'-end), which is strongly preferred, and a parallel orientation. The melting temperature of PNA-nucleic acid duplexes is largely independent of the salt concentration. Although PNA prefers a P-form conformation, it adapts the structure of its nucleic acid partner. An RNA-PNA duplex is in A-form conformation, DNA-PNA in a B-like structure. A further benefit of PNA is, that PNA-RNA hybrids are not affected by degradation due to RNase H, in contrast to RNA-DNA duplexes.[155] This fact is of high importance in terms of studies *in vivo.*

Following the convention, PNAs are depicted like peptides from the *N*-terminus at first position and *C*-terminus at right.[152]

3.4.1 Principle of Solid Phase PNA Synthesis

Another advantage of PNA is its straightforward synthesis, which is based on the well-established solid phase peptide synthesis (SPPS). The robustness and flexibility of this synthesis allows an individual tunable design.[152]

The general principle of solid phase PNA synthesis is shown in Scheme 6.[156]



Scheme 6: Solid phase PNA synthesis. Free amine groups (red) react with activated esters to form a stable amine bond. The desired PNA sequence is synthesized by subsequent coupling of PNA monomers at a polymer resin support (brown) with Fmoc deprotection steps in between and final cleavage from the resin.

Usually, the PNA strand is synthesized from the *C*- to the *N*-terminus. The solid support on which the reactions are carried out is a polymer resin with adjustable functionalities. Starting from primary amines at the resin, activated PNA monomers are coupled to form stable amide bonds. Every monomer is Fmoc (fluorenylmethyloxycarbonyl) protected at its amine function to ensure that only the amine groups on the resin can react. Additionally, the particular bases are Bhoc (benzhydryloxycarbonyl) protected at their primary amine side chain functionalities. This strategy allows to selectively deprotect the backbone amine function for coupling without side reactions at the nucleobases. Monomers are coupled via active ester formation with HATU (1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexa-fluorophosphate).[157] Completion of each coupling step can be monitored by the Kaiser test.[158,159] This test is based on a Schiff base reaction with a color change to blue if free amines are present. After Fmoc deprotection the next monomer can be coupled until the sequence is completed. The final product can be cleaved from the resin under acidic conditions, which simultaneously remove the Bhoc-protecting groups on the nucleobases.

3.4.2 Applications of PNA

With the above described useful properties PNA is already in use for a variety of applications. PNAs can be used in drug-discovery with the potential as gene-targeted drugs,[160–162] antisense agents,[163] genetic diagnostics,[164] molecular recognition,[165] DNA detection, and therapeutics.[152] Modified PNAs were investigated for template-accelerated chemical reactions for the detection of mRNA's expression level or microRNAs *in vivo*.[166] PNA anti-miRs (oligonucleotide inhibitors complementary to miRNAs) together with intrinsically fluorescent PNA analogues are used for studies upon inhibition of miRNAs as potential therapeutics.[167] Furthermore, fluorescent PNA-based probes have been used to explore the splicing of an mRNA molecule by smFRET. To this end, the flanking splice sites were labeled by hybridization of dye-carrying PNAs and the spatial separation of the dyes was detected before and after splicing of the RNA.[168] PNA has many more applications in the field of chemistry, biology, and medicine. The studies described herein are focused on using PNAs as hybridization probes for fluorescence labeling to investigate the folding of large RNAs by smFRET.

3.5 Aim of the Chapter

As mentioned before (Chapter 3.3), modifications within the RNA structure for an indirect labeling with DNA hybridization probes can disturb the structure and thus the functionality of the RNA. To overcome these drawbacks and improve the labeling strategy, we envisaged the use of the non-natural nucleic acid analogue Peptide Nucleic Acid (PNA) instead of DNA oligomers for efficient fluorescence labeling. A significant advantage of PNA over DNA is its higher mismatch sensitivity which makes it an extremely selective labeling agent.[153,154] Furthermore, PNA is known to bind much stronger to RNA than DNA does due to its uncharged backbone (Chapter 3.4).[154] Consequently, the PNA probes can be drastically shortened compared to the analogues DNA oligomers, and smaller or even no modifications of the RNA are necessary for the hybridization. Ideally, hybridization to the natural loops within the RNA structure would suffice. This leads to more flexibility in choosing the labeling position within the target construct and thereby less disturbance of the folding.

In the following chapter, the investigation of the optimal label design together with a systematic comparison of the commonly used RNA-DNA labeling scheme with the complementary RNA-PNA system is described. To this end, new fluorescent PNA derivatives containing the FRET dyes Cy3 or Cy5 are synthesized and characterized including photophysical properties. Furthermore, the optimal length of the PNA sequence, the position of the dye, the labeling efficiency, and potential interference of PNA with the RNA structure, dynamics and activity is determined with various methods as a proof of concept study on the RNA model system *Sc*.ai5γ.

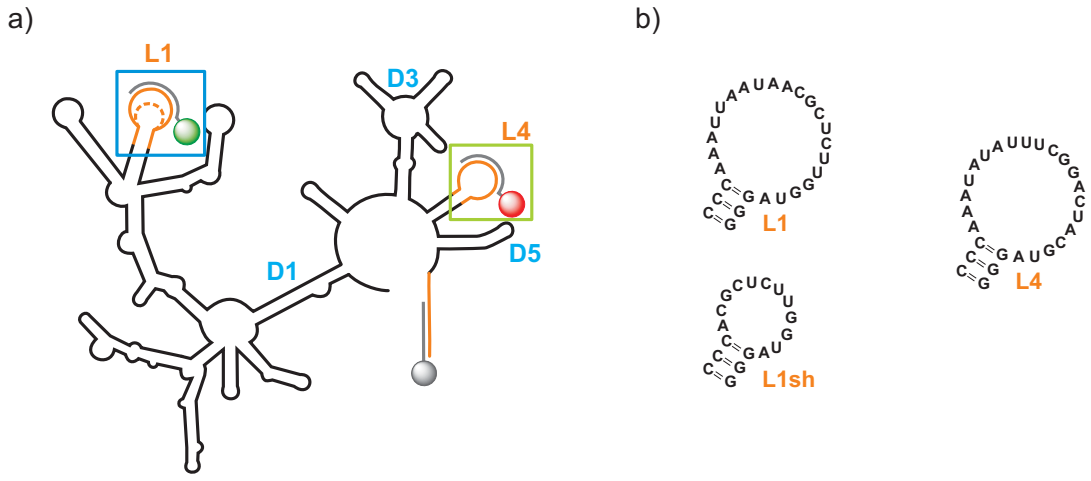


Figure 50: Fluorescence labeling strategy for RNA using a model system. a) Schematic drawing of D135-L14 (solid line) and D135-L1shL4 with shortened L1 (dotted line), and hybridized oligonucleotide labels Cy3 (green), Cy5 (red), and biotin (grey) for immobilization. Modifications of D135 are depicted in orange (L1 = modified loop in domain 1, L4 = modified loop in domain 4). b) Sequence of artificial hairpin loops L1, L1sh (blue box in a) and L4 (green box in a) located in the RNA construct D135-L14 or D135-L1shL4.

Therefore, a truncated version of the mitochondrial group II intron ai5γ derived from *S. cerevisiae* (D135-L14) was chosen, because it is a well-characterized large RNA molecule (> 600 nts) with enzymatic activity (Figure 47, Figure 50, Chapter 3.2.3).[42] To demonstrate that the required structural modification of RNA for hybridization are minimized when using the shorter PNA labels, one or both of the artificially introduced loops (L1 and L4) of the D135-L14 construct are shortened by 9 nucleotides each, yielding D135-L14sh and D135-L1shL4 (Figure 50). This labeling strategy thus allows investigating the folding behavior of large functional RNAs closer to the natural sequence. Finally, the applicability is proven by use of smFRET.

After investigations of the PNA labeling scheme and systematic comparison using a model system, the labeling scheme is applied to a new RNA system, the *btuB* riboswitch of *E. coli*.

3.6 Results

The following chapter is based on the manuscript accepted for publication:

“Internal Labeling Strategy of Large RNAs with Minimal Perturbation Using Fluorescent PNA”

by Anita G. Schmitz, Susann Zelger-Paulus, Gilles Gasser, and Roland K. O. Sigel in **ChemBioChem**, 2015.[169]

3.6.1 Synthesis of PNA Derivatives

The ribozyme construct D135-L14 was designed to bind DNA-dye oligomers functioning as indirect fluorescent labels (Chapter 3.2.3). To ensure a stable hybridization to the RNA the DNA oligomers (**DNA1-Cy3**, **DNA2-Cy5**) have a length of 18 nts (Figure 51).

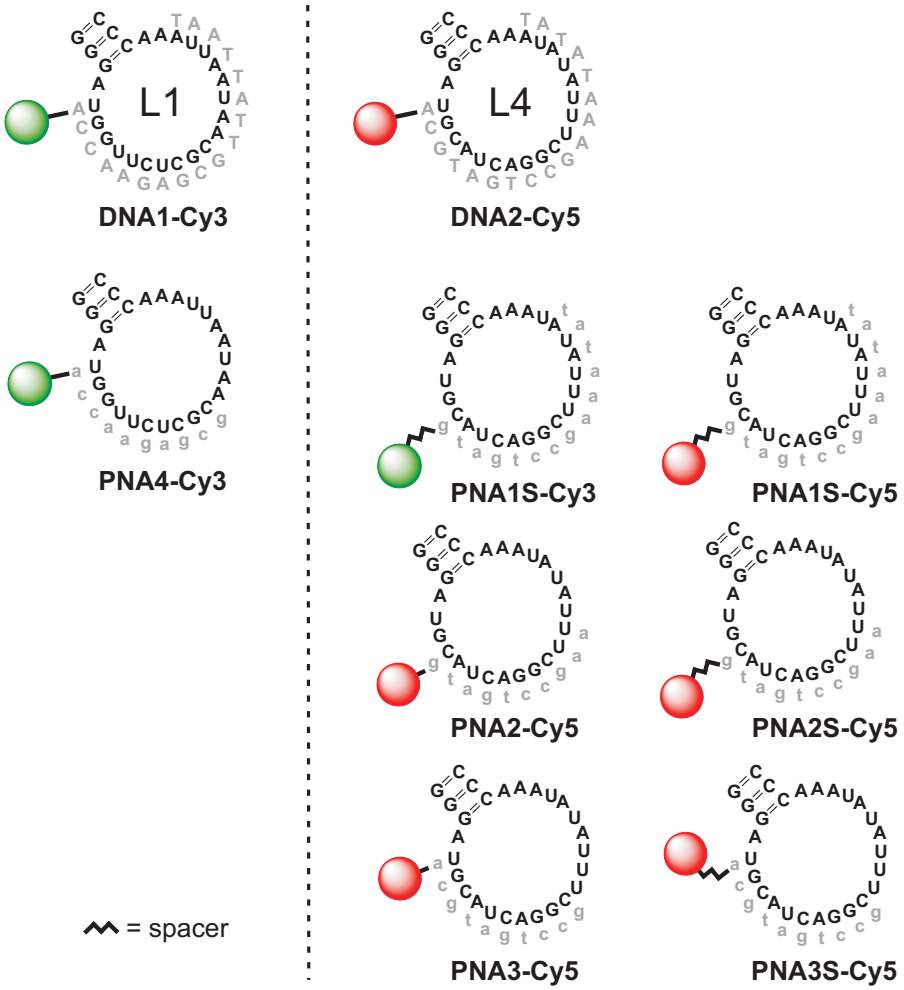


Figure 51: DNA-Cy and PNA-Cy derivatives complementary to loops L1 and L4 in D135-L14. Design and nomenclature of synthesized PNA derivatives carrying Sulfo-Cy3 (green) or Sulfo-Cy5 (red) in comparison to corresponding DNA derivatives. DNA and PNA sequences (grey) are complementary to L1 or L4 located in the D135-L14 RNA (black). PNA derivatives with an additional spacer group are indicated by an ‘S’.

Based on this, different PNA sequences were designed to carry the fluorescent cyanine dyes (Cy) Sulfo-Cy3 or Sulfo-Cy5 (FRET pair), and to be hybridized to the loops L1 and L4 in D135-L14. Due to the strong binding affinity of PNA to RNA (Chapter 3.4) the derivatives are shorter (10-mers and 14-mers) than the corresponding DNA (18-mer) oligonucleotides (Figure 51).

The PNA derivatives were prepared by use of solid phase synthesis (see Chapter 3.4.1).[170] To these PNA sequences the cyanine dyes Sulfo-Cy3 or Sulfo-Cy5 were covalently attached at the *N*-terminus, which is analogous to the 5' end in DNA, with or without a spacer group (2-(2-(amino)ethoxy)ethoxy]acetate, AEEA) in between. An additional *C*-terminal lysine residue was used to increase the solubility. All derivatives in this chapter contain the sulfonated cyanine dyes, which are abbreviated with Cy3 or Cy5 for simplicity. These negatively charged derivatives are commonly used in nucleic acids studies (Chapter 1.2.2) due to their low tendency to intercalate.

The overall structural principle of the PNA-Cy derivatives is shown in Figure 52 from *N*- to *C*-terminus containing a dye, a spacer, a PNA sequence built from monomers and lysine.

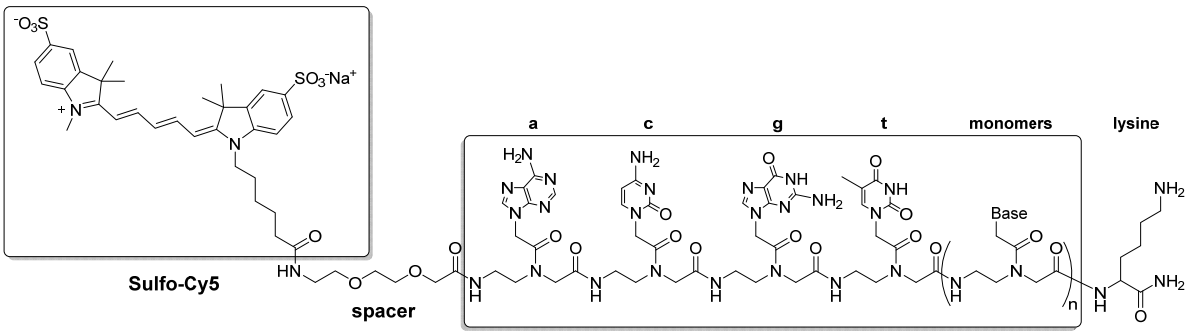
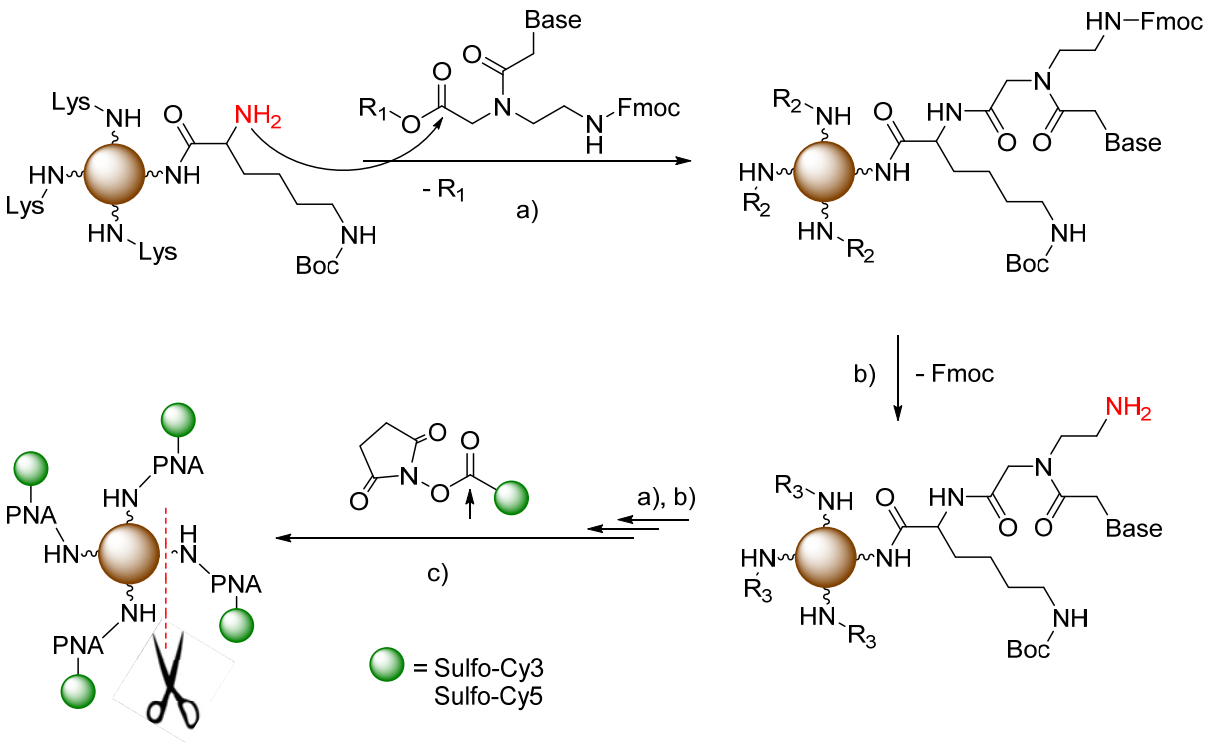


Figure 52: Structure of a PNA-Cy5 derivative. Special subunits are highlighted in boxes. The derivatives consist of a cyanine dye, optionally a spacer group, various monomers coupled to give the desired sequence, and lysine for increased solubility. PNA bases are denoted in small letters from *N*- to *C*-terminus.

The synthesis starts from a modified polymer resin with bound lysines that are Boc protected at the ε-amino function and Fmoc protected at the α-amino function (Scheme 7). Fmoc is cleaved from the lysines with piperidine in DMF to give free primary amine groups. The PNA monomers are activated with HATU in the presence of *N*,*N*-diisopropylethylamine (DIPEA) and 2,6-lutidine, and coupled to the primary amines. In a next step, Fmoc is removed from the terminal backbone amine groups to allow the next coupling step. The dyes Sulfo-Cy3 or Sulfo-Cy5 are used as NHS esters and coupled to the PNA sequence in the presence of DIPEA. Finally, Boc deprotection and cleavage of the completed PNA sequence from the resin is done under acidic conditions by use of TFA.

Resins commonly used in SPPS are known to absorb activated fluorophores leading to a decreased labeling yield.[171] Hence, in an alternative approach the resin with the **PNA1S** sequence was saturated with Kiton Red, an inexpensive non-activated fluorophore, prior to

the addition of Sulfo-Cy3. In the present case, however, the procedure did not increase the yield significantly.



Scheme 7: Solid phase synthesis of PNA-Cy derivatives. The synthesis starts from a lysine-modified resin (brown). a) Subsequent coupling of Fmoc protected PNA monomers. b) Fmoc cleavage with piperidine. Repetition of a) and b) until the sequence is completed. c) Dye coupling in the presence of DIPEA in DMF yielding product PNA-Cy which is cleaved from the resin with simultaneous Boc deprotection of the nucleobases.

The 14-mer **PNA1S-Cy** derivative was synthesized in two variants, once carrying the donor Sulfo-Cy3 and once the acceptor Sulfo-Cy5 (see Figure 51). This offers the possibility to exchange the donor and acceptor positions within the RNA construct. The investigation of different labeling positions can help excluding artifacts in smFRET measurements. Ideally both positions give the same results. To minimize the modifications in the RNA structure needed for label hybridization, the PNA oligomers should be as short as possible while keeping sequence specificity and duplex stability. Therefore different PNA sequence lengths were investigated. **PNA2** and **PNA3** derivatives (Figure 51) are designed to be 10-mers, which are thus 4 nucleobases shorter than **PNA1S** and 8 nucleobases shorter than the DNA oligomers **DNA1-Cy3** and **DNA2-Cy5**. To form stable PNA-RNA hybrids and reduce possible disruption of the stem-loop structure, the label should not be in close distance to the RNA duplex. To investigate the potential influence of the final position of the dye in PNA-RNA hybrids **PNA2-Cy5** and **PNA3-Cy5** derivatives differ in their sequence. **PNA3-Cy5** ends at the same positions as **DNA2-Cy5**, whereas **PNA3-Cy5** is shifted by two bases. Furthermore, the influence of a spacer group between the oligonucleotides and the dye was investigated. In general, commercially available dye-labeled DNA oligonucleotides can vary in the spacer

group between the DNA and the dye. For this reason, derivatives of the same sequence were prepared, once with the dye directly linked to the PNA (**PNA2-Cy5**/**PNA3-Cy5**) or with a spacer group in between (**PNA2S-Cy5**/**PNA3S-Cy5**). The spacer (Figure 52) has a similar length as the C6-linker in the DNA probes used in this study.
Each PNA derivative was characterized by UV/Vis, HPLC, ESI-MS and MALDI-MS (Chapter 3.6.2).

3.6.2 Characterization of PNA Derivatives

UV/Vis

In UV/Vis spectra of the PNA-dye derivatives the characteristic absorption maxima of PNA at 260 nm and Sulfo-Cy3 at 550 nm or Sulfo-Cy5 at 650 nm were observed. The fluorescence emission of Sulfo-Cy3 and Sulfo-Cy5 is similar to the free dyes at around 570 nm and 670 nm (Figure 53).

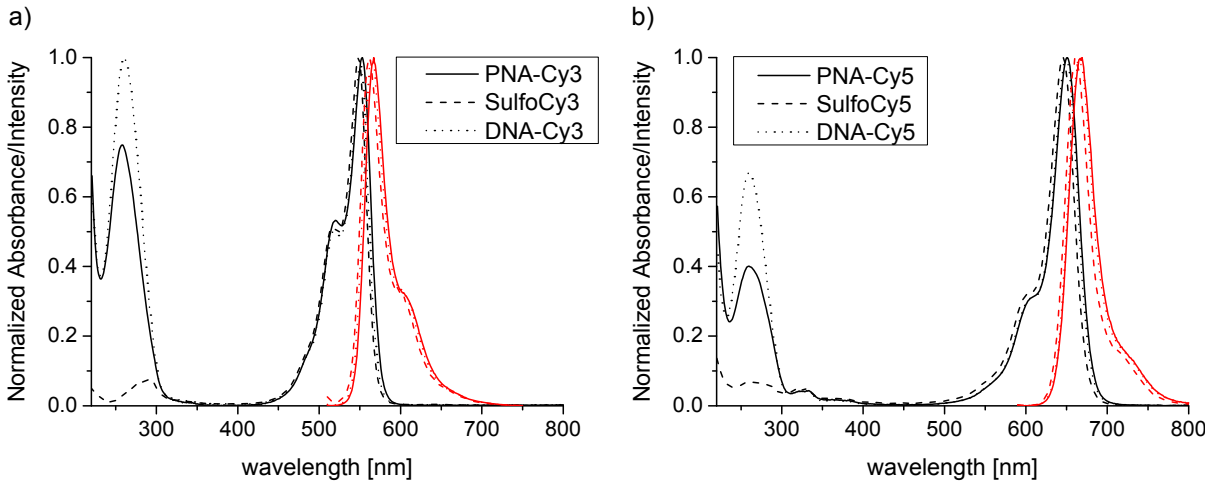


Figure 53: Absorbance and emission spectra of PNA-Cy derivatives in comparison to DNA-Cy derivatives and free Cy dyes. Absorbance (black) and emission spectra (red) in aqueous solution of a) **PNA4-Cy3** (solid line), **DNA1-Cy3** (dotted line) and Sulfo-Cy3 (dashed line), and b) **PNA3-Cy5** (solid line), **DNA2-Cy5** (dotted line) and Sulfo-Cy5 (dashed line). Sulfo-Cy3 is excited at 500 nm, Sulfo-Cy5 at 580 nm.

The results of the PNA-Cy derivatives are consistent with the ones of the corresponding DNA-dye derivatives.

HPLC

In HPLC experiments the respective product peaks of the PNA-dye derivatives were detected at 260 nm (absorbance PNA) and 550 nm (absorbance Sulfo-Cy3) or 650 nm (absorbance Sulfo-Cy5), and emission at 570 nm (Sulfo-Cy3) or 670 nm (Sulfo-Cy5). A

comparison of the chromatograms of **PNA1S-Cy3** at different wavelength is shown in Figure 54.

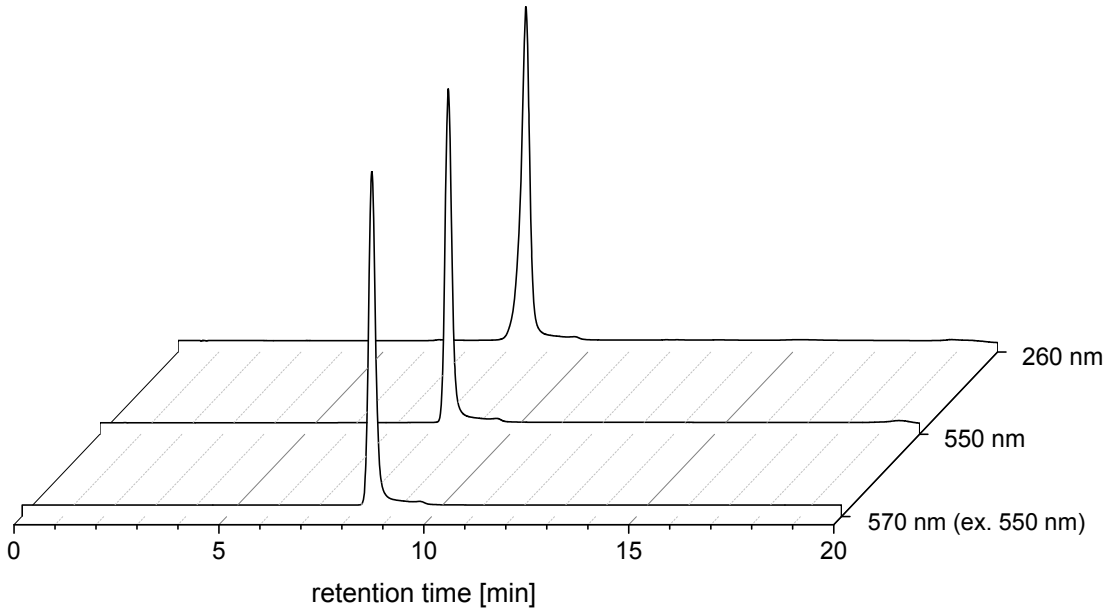


Figure 54: HPLC chromatograms of PNA1S-Cy3 at different wavelength. At 260 nm the absorption of the PNA is detected, at 550 nm the absorption of the Sulfo-Cy3. The Sulfo-Cy3 emission is detected at 570 nm upon excitation at 550 nm. The intensities are normalized.

Mass Spectrometry

In ESI-MS and MALDI-MS experiments product peaks of the different PNA-Cy derivatives were observed. The detection of differently charged species due to multiple protonation is characteristic for PNA oligomers (Figure 55 and Figure 56).

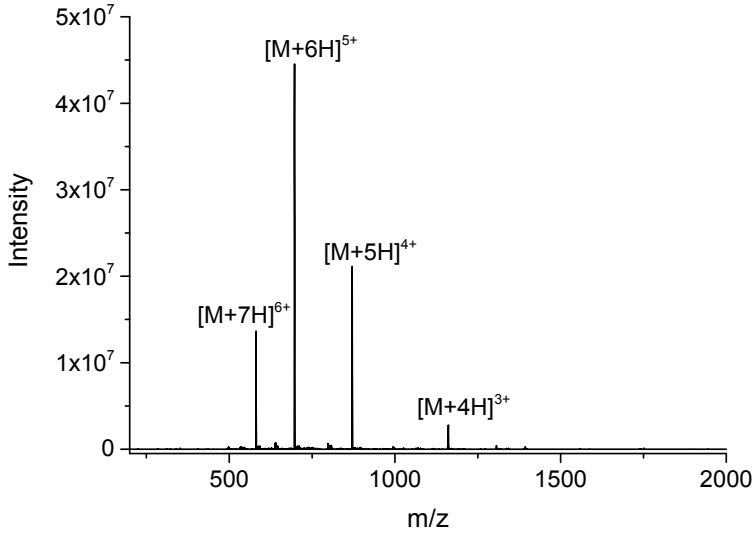


Figure 55: ESI-MS spectrum of PNA3-Cy5. Product peaks of **PNA3-Cy5**. Multiple protonated species are characteristic for PNA oligomers.

The ESI-MS spectrum of **PNA3-Cy5** (Figure 55) reveals the product peaks of the four to seven times protonated species.

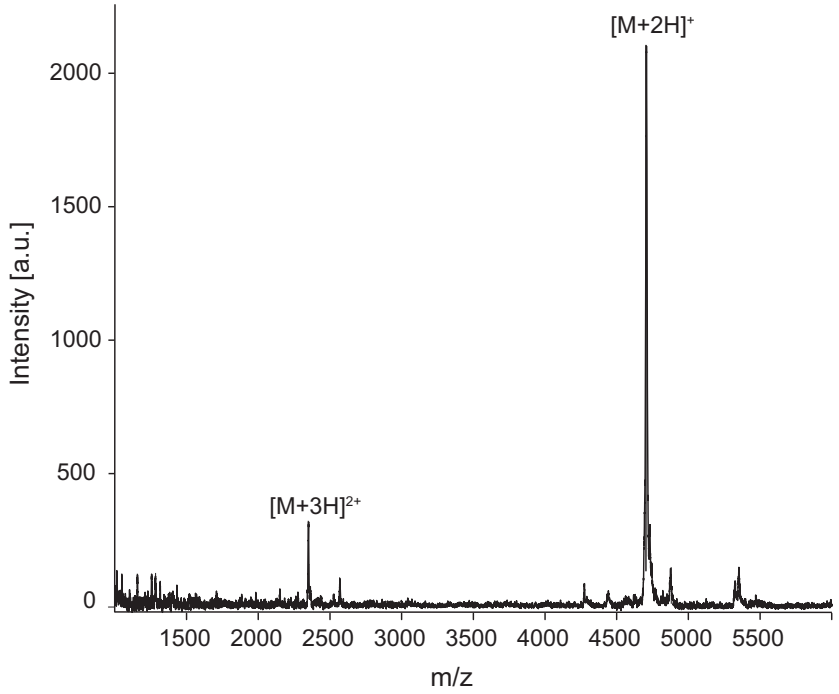


Figure 56: MALDI-MS spectrum of PNA1S-Cy3. Product peaks of **PNA1S-Cy3**.

In the MALDI-MS spectrum of **PNA1S-Cy3** (Figure 56) the product peaks are detected as doubly or triply protonated species.

3.6.3 Fluorescence Quantum Yields and Steady-State Anisotropy

The PNA labels are designed for fluorescence applications, especially FRET. Provided that the photophysical properties of the probes are unaffected by the environment, variations in the measured FRET efficiency can then be interpreted in the line of conformational changes of the RNA (Chapter 1.2). The fluorescence quantum yields of Sulfo-Cy3 and Sulfo-Cy5 attached to DNA and PNA oligonucleotides were therefore determined (Table 3). Quantum yields of **PNA4-Cy3** (0.26 ± 0.03) in comparison to **DNA1-Cy3** (0.26 ± 0.02), and **PNA3-Cy5** (0.32 ± 0.01) in comparison to **DNA2-Cy5** (0.35 ± 0.01) revealed no differences between both labeling systems. DNA is known to enhance the fluorescence of attached cyanine dyes in comparison to free cyanine dyes. This enhancement is based on the reduction of the *trans-cis* isomerization from the first excited state to a ground state non-fluorescent photoisomer (Chapter 1.2.2).[21] The extent of the effect depends on the neighboring nucleobases.[20,28] In the herein presented studies the same enhancement effect was observed for the PNA hybrids. The effect is slightly reduced when the spacer group is present as the dye is further apart from the neighboring nucleobases.

Additionally, steady-state fluorescence anisotropy was measured to determine the changes of rotational freedom of the dye upon attachment to DNA or PNA. All derivatives showed similar anisotropies of around 0.20 (Table 3). Sulfo-Cy5 attached to PNA or DNA revealed a higher anisotropy value than the free dye. A similar behavior was observed for B_{12}-Cy5/Sulfo-Cy5 derivatives (Chapter 2.5.2).

These results demonstrate that PNA-Cy derivatives have similar photophysical properties as the previously used DNA-Cy oligonucleotides.

Table 3*:* Fluorescence quantum yields (Φ_s) and static fluorescence anisotropy values (r) of free Sulfo-Cy3 (excited at 500 nm), free Sulfo-Cy5 (excited at 580 nm) or attached to DNA or PNA oligonucleotides at 22 °C in a) *Reaction Buffer* with 100 mM $MgCl_2$ and b) water. The values given correspond to the arithmetic mean of at least three independent measurements with errors corresponding to one standard deviation.

Compound	Φ_s (a)	Φ_s (b)	r
Sulfo-Cy3	0.071 ± 0.002	0.060 ± 0.002[a]	0.243 ± 0.005
PNA1S-Cy3	-	0.24 ± 0.02	0.22 ± 0.01
PNA4-Cy3	0.26 ± 0.03	0.25 ± 0.02	0.22 ± 0.01
DNA1-Cy3	0.26 ± 0.02	0.15 ± 0.02	0.23 ± 0.01
Sulfo-Cy5	0.23 ± 0.01	0.203 ± 0.006[b]	0.131 ± 0.009
PNA1S-Cy5	0.30 ± 0.02	0.32 ± 0.02	0.204 ± 0.005
PNA2-Cy5	0.35 ± 0.02	0.278 ± 0.004	0.193 ± 0.008
PNA2S-Cy5	0.30 ± 0.02	0.26 ± 0.01	0.183 ± 0.005
PNA3-Cy5	0.32 ± 0.01	0.34 ± 0.02	0.186 ± 0.010
PNA3S-Cy5	0.29 ± 0.01	0.273 ± 0.001	0.188 ± 0.004
DNA2-Cy5	0.35 ± 0.01	0.28 ± 0.02	0.217 ± 0.007

[a] Quantum yield provided by supplier Φ= 0.1

[b] Quantum yield provided by supplier Φ= 0.2

The quantum yields of the PNA-Cy derivatives are mainly independent from the ionic strength of the solution.

3.6.4 UV Melting Experiments

To assess the duplex stability of RNA-DNA in comparison to RNA-PNA, UV melting experiments were carried out with the hybrids formed between dye-labeled PNA or DNA oligonucleotides and the corresponding short complementary RNA sequences.[172]

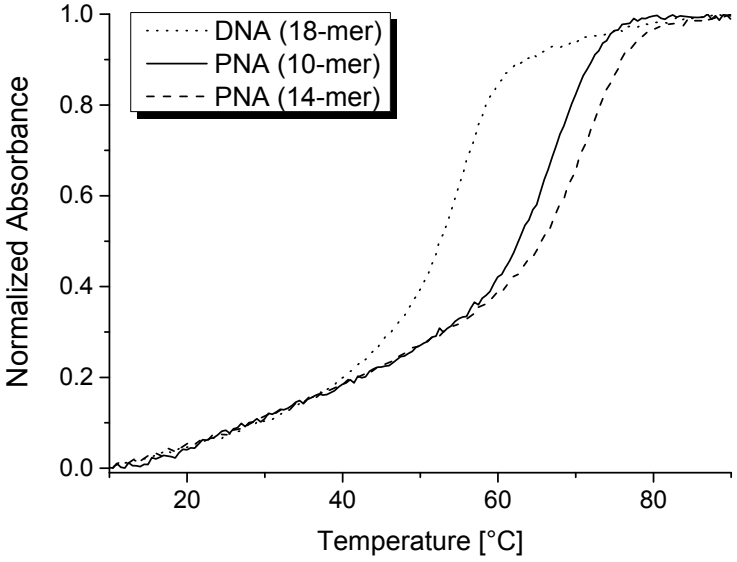


Figure 57: Melting curves of DNA-RNA and PNA-RNA hybrids of different lengths. The melting temperature of DNA2-Cy5 (dotted line) is lower than the one of **PNA1S-Cy5** (dashed line) and **PNA3-Cy5** (solid line).

A comparison of RNA hybrids with a 18 nts long DNA-Cy oligomer, a 14 nts PNA-Cy oligomer and a 10 nts long PNA-cy oligomer is shown in Figure 57.

Furthermore, different synthesized PNA-Cy derivatives were used to study the influence of a spacer group between PNA and dye, as well as differences in the distance from the label to the stem in the D135-L14 structure (Chapter 3.6.1). All melting temperatures (T_m) of the RNA-PNA hybrids are at least 10 °C higher in comparison to the RNA-DNA duplexes, although the PNA sequences are 4-8 nucleobases shorter than the corresponding DNA oligomers (Table 4). This shows that the binding strength of the shortened PNA oligomers to the RNA is even higher than the one of the longer DNA oligomers.

Table 4*:* Melting temperatures of RNA-PNA and RNA-DNA hybrids with [RNA] = [DNA] = [PNA] = 2 µM in *Reaction Buffer*. The given T_m values are the average of at least six measurements.

PNA/DNA	Sequence[a]	T_m [°C]
PNA1S-Cy3	Cy3-spacer-gtagtccgaaatat-Lys-NH_2	70 ± 1
PNA1S-Cy5	Cy5-spacer-gtagtccgaaatat-Lys-NH_2	72 ± 1
PNA2-Cy5	Cy5-gtagtccgaa-Lys-NH_2	69 ± 2
PNA2S-Cy5	Cy5-spacer- gtagtccgaa-Lys-NH_2	69 ± 1
PNA3-Cy5	Cy5-acgtagtccg-Lys-NH_2	67 ± 1
PNA3S-Cy5	Cy5-spacer-acgtagtccg-Lys-NH_2	66 ± 1
PNA4-Cy3	Cy3-accaagagcg-Lys-NH_2	69 ± 1
DNA1-Cy3	5'-Cy3-ACCAAGAGCGTTATTAAT-3’	57 ± 1
DNA2-Cy5	5'-Cy5-ACGTAGTCCGAAATATAT-3’	56 ± 1

[a] spacer = [2-(2-(amino)ethoxy)ethoxy]acetate

A theoretically possible destabilization of the duplex by the close proximity of the dye to the hybrid, as it is the case in **PNA2-Cy5**, **PNA3-Cy5** and **PNA4-Cy3**, was not observed. This indicates as well that the spacer group between the PNA and the dye does not influence the duplex stability.

3.6.5 New Constructs D135L1sh4 and D135L14sh

PNA with its stronger binding affinity to RNA gives the opportunity to use shorter oligonucleotide labels of only 10 nts (see above) and hence, smaller modifications of the RNA are needed for the hybridization, leading to a RNA sequence closer to the natural one. To confirm this, the D135-L14 construct was modified to be more similar to its wild type structure by the deletion of 9 bp (5'-AAUUAAUAA-3') in loop L1 yielding the RNA construct D135-L1shL4 (Figure 58).

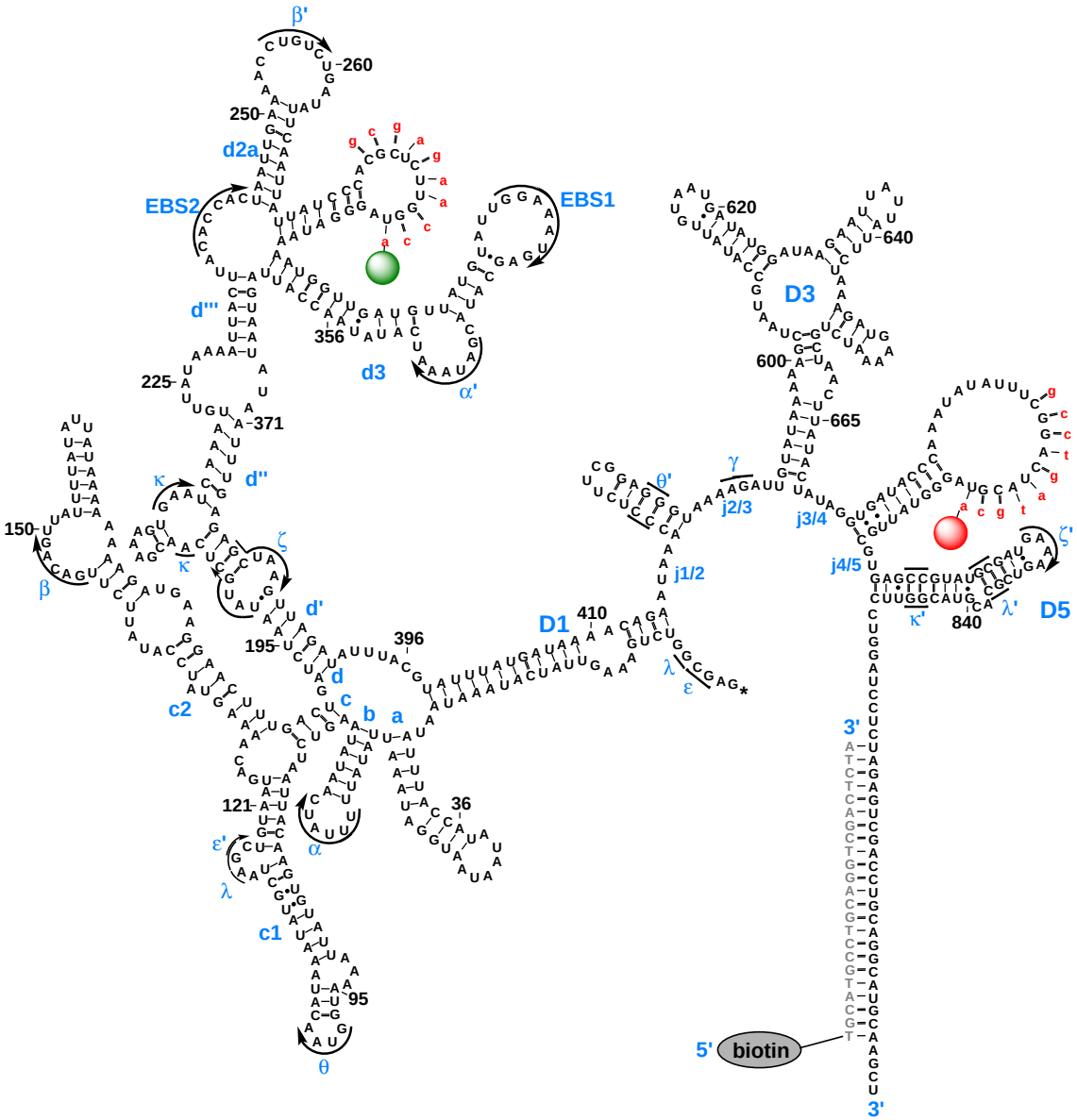


Figure 58: Secondary structure of D135-L1shL4 with PNA-dye labels. PNA oligonucleotides (red) carrying Sulfo-Cy3 (green) or Sulfo-Cy5 (red), and DNA-biotin (grey) are hybridized to D135-L1shL4 RNA. The numbering corresponds to the wild-type *Sc*.ai5γ sequence and the tertiary interactions are indicated in Greek letters.

Additional deletion of 9 bp (5'-AAUAUAUUU-3') in loop L4 yielded the RNA construct D135-L14sh (Figure 59).

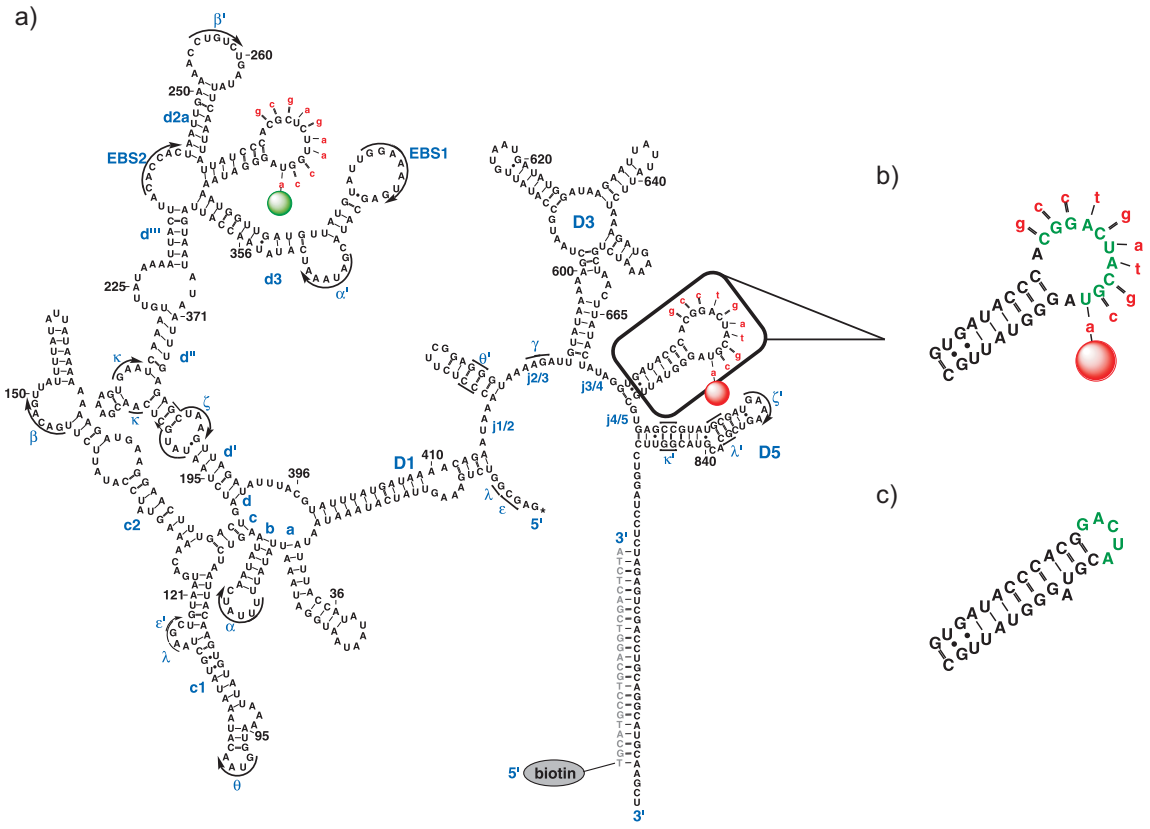


Figure 59: Secondary structure of D135-L14sh with PNA-dye labels. a) D135-L14sh. b) Close-up of desired structure of shortened loop L4 for binding of PNA-Cy labels. c) Predicted secondary structure with minimum free energy of loop L4 (-14.40 kcal/mol).

However, fluorescence native gel experiments revealed that the complementary PNA derivatives were not binding to the new loop L4 (data shown in Chapter 3.6.6). Secondary structure prediction of the shortened loop L4 sequence did not confirm the presence of 12 unpaired nucleotides as anticipated, but suggests that a stable G-C rich stem with a pentaloop (Figure 59) is formed.[173] This structure does not provide a sufficient hybridization platform for the PNA to bind. D135-L14sh was thus not used for smFRET studies.

3.6.6 Fluorescence Native Gels

To visualize the binding efficiency and specificity of the fluorescently labeled PNA oligomers to RNA in comparison to DNA oligomers native polyacrylamide gel electrophoresis assays were performed with D135-L14, D135-L14sh and D135-L1shL4 (Figure 60). The conditions were chosen to be optimal for folding of the ribozyme *in vitro* (Chapter 3.2.3) and are similar to the ones used for single-molecule FRET studies.[42,145] In a first step the labeling procedure was optimized. Due to the high binding affinity of PNA to RNA, PNA can be hybridized at 42 °C, whereas DNA has to be annealed under denaturing conditions at 90 °C. High

temperatures can lead to dye deactivation, RNA degradation as well as RNA aggregation. As this can be avoided with PNA, it is one reason for an increased labeling yield of PNA compared to DNA. Furthermore, the labeling procedure is faster without the denaturation step.

Binding of **PNA4-Cy3** and **PNA3-Cy5**, individually (Figure 60 a, lanes 8 and 9, respectively) and simultaneously (lane 10), to D135-L14 RNA was investigated in comparison to **DNA1-Cy3**/**DNA2-Cy5** (lanes 3-5, 11). PNA-Cy3/Cy5 oligonucleotides which carry a charge only at the dyes hardly enter the gel. The yellow bands indicate the presence of both, Cy3 and Cy5 labels bound to the RNA (lanes 5, 10, 11). The three main bands (A, B, C) from top to bottom are aggregated/unfolded, intermediate and compact species of the RNA. Controls in lane 1, 2 and 6, 7 are the DNA (D) or PNA oligonucleotides, respectively without RNA.

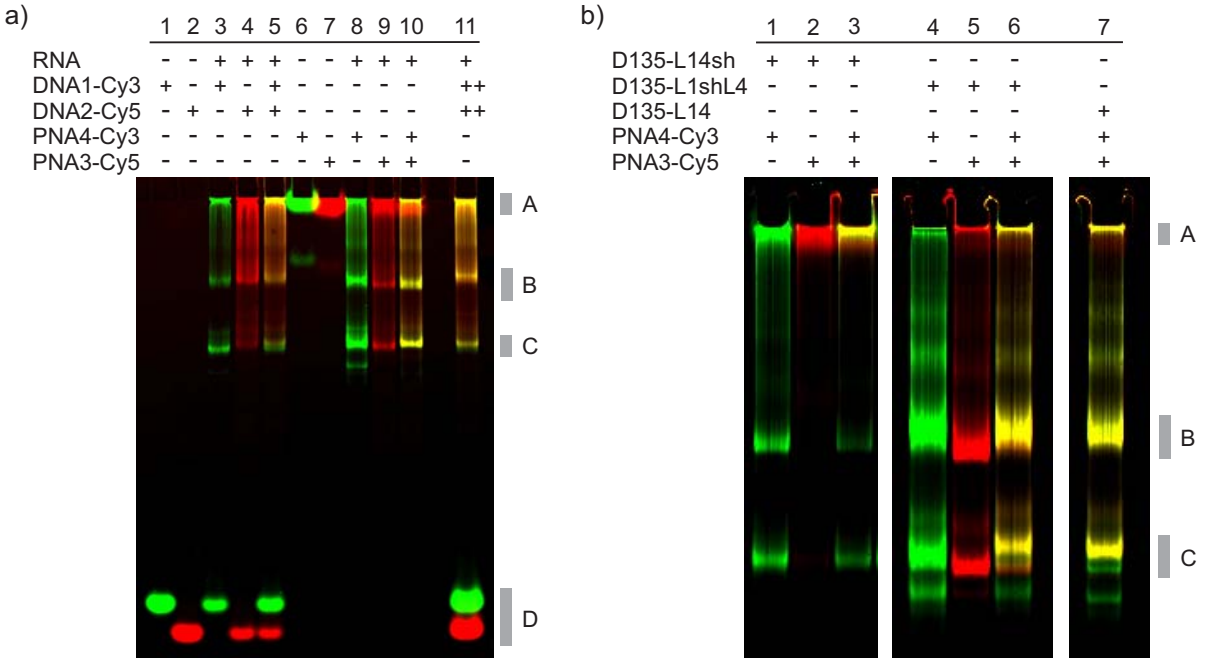


Figure 60: Fluorescence native gels of D135-L14sh, D135-L1shL4, and D145-L14 RNA with DNA-dye or PNA-dye labels. Binding of DNA or PNA oligonucleotides carrying the label Sulfo-Cy3 (green) or Sulfo-Cy5 (red) to a) D135-L14 and b) D135-L14sh (lanes 1-3), D135-L1shL4 (lanes 4-6), and D135-L14 (lane 7) detected by fluorescence on native gels. Co-localized fluorescence emission of Sulfo-Cy3 and Sulfo-Cy5 appears in yellow. Grey bars indicate (A) unbound PNA oligomers or labeled RNA aggregates, (B) RNA intermediates, (C) compact RNA, and (D) unbound DNA-Cy3/Cy5 oligomers. Composition of samples depicted as present (+), absent (-) or in excess (++).

The superior labeling efficiency of PNA over DNA is reflected in a much higher fraction of labeled RNA in case of PNA, a large fraction of unbound DNA (lowest bands) and almost no unbound PNA. With respect to FRET measurements, the higher fraction of labeled molecules leads to a reduced fluorescence background. Furthermore, less PNA is necessary for efficient labeling compared to DNA, since PNA showed the highest efficiencies at a ratio of

1:1 RNA/PNA (Figure 61), whereas the efficiency of DNA is lower even in twofold excess (Figure 60 a, lane 11).

Fluorescence native gels of D135-L1shL4 revealed as well a very good binding of the PNA to the loops L1sh and L4 (Figure 61 b). This construct containing the shortened loop L1sh cannot be labeled with DNA probes because the modified loop is too small to form a sufficiently stable RNA-DNA duplex.

In experiments with D135-L14sh and PNA labels a binding of PNA to the shortened loop L1sh war observed (Figure 60 b). However, no labeling at loop L4sh was detected due to the unfavorable loop structure (Chapter 3.6.5).

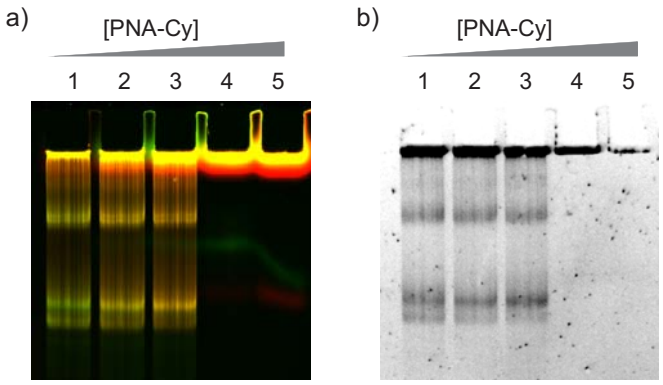


Figure 61: Fluorescence native gels of D135-L14 RNA with PNA-dye labels in different ratios. Increasing concentrations of PNA-Cy from lanes 1-5 with ratios of RNA:PNA 2:1, 1:1, 1:1.2, 1:1.5, 1:2. a) Binding of **PNA4-Cy3** (green) and **PNA3-Cy5** (red) to D135-L14 detected by fluorescence on native gels. Co-localized fluorescence emission of Sulfo-Cy3 and Sulfo-Cy5 appears in yellow. b) EtBr staining of RNA bands detected by UV shadowing at 260 nm.

In further studies with increasing concentrations of PNA hybridized to D135-L14 no folded RNA band was detected at ratios of RNA:PNA of 1:1.5 and higher (Figure 61). However, no RNA degradation bands were observed (Figure 61 b). This suggests that high concentrations of uncharged PNA could lead to an aggregation of the RNA. Due to its size aggregated RNA is hardly entering the gel and is consequently not fully detected.

3.6.7 Single-Turnover Cleavage Assays

It is of great interest to study the folding of functional RNAs. Therefore, the labeling method for RNA should not reduce the activity of the RNA. The effect of modifications within the intron sequence can be directly evaluated by measuring its catalytic activity. In a so-called cleavage assay mimicking the first step of splicing, an external substrate (17/7) is cleaved by group II intron under single turn-over conditions.[138] The substrate contains the last 17 nucleotides of the 5'-exon sequence including both intron binding sites IBS1 and IBS2 and the first 7 nucleotides of the intron and thus the 5' splice site (Figure 62). The cleavage rate of the substrate reflects the activity of the ribozyme.

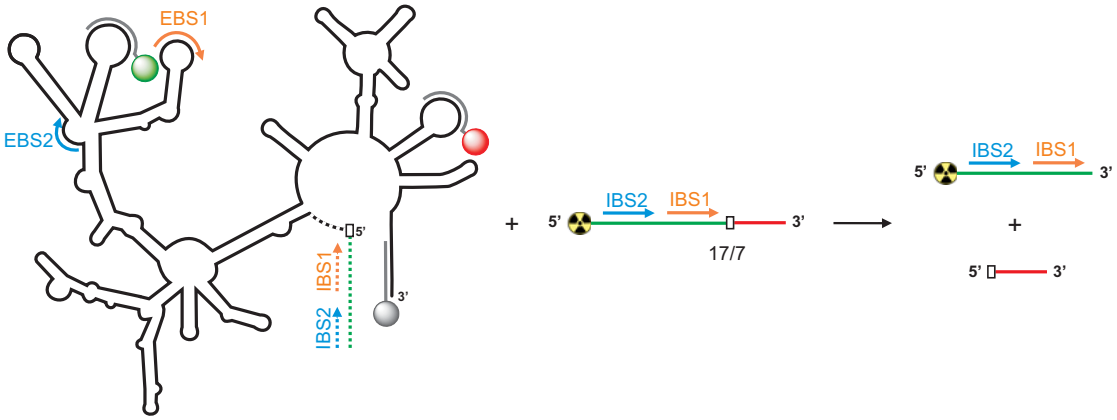


Figure 62: Principle of ribozyme cleavage assay. The ribozyme RNA (schematic drawing of D135-L14 in black with hybridized DNA or PNA labels in grey) cleaves the exonic substrate 17/7, which contains the intron binding sites IBS1/2 into the fragment 17 and 7. The location of the substrate in the natural sequence is depicted in dotted lines. For visualization purpose, the substrate is [^{32}P]-5'-end labeled (radioactive symbol).

To avoid the presence of unlabeled ribozyme in the experiment, DNA-dye oligomers were added in a 6-fold excess. PNA-dye oligomers were added in a 1:1 ratio of PNA/RNA, as this was found to be the labeling condition with the highest efficiency (Chapter 3.6.6). The activity of the different *in vitro* transcribed ribozyme constructs D135, D135-L14, D135-L14sh and D135-L1shL4 was investigated by single-turnover cleavage experiments (STO). The 17/7 substrate was [^{32}P]-5'-end labeled for a sensitive detection, and the cleavage of the substrate by the ribozyme with or without hybridized DNA or PNA was analyzed (Figure 63).

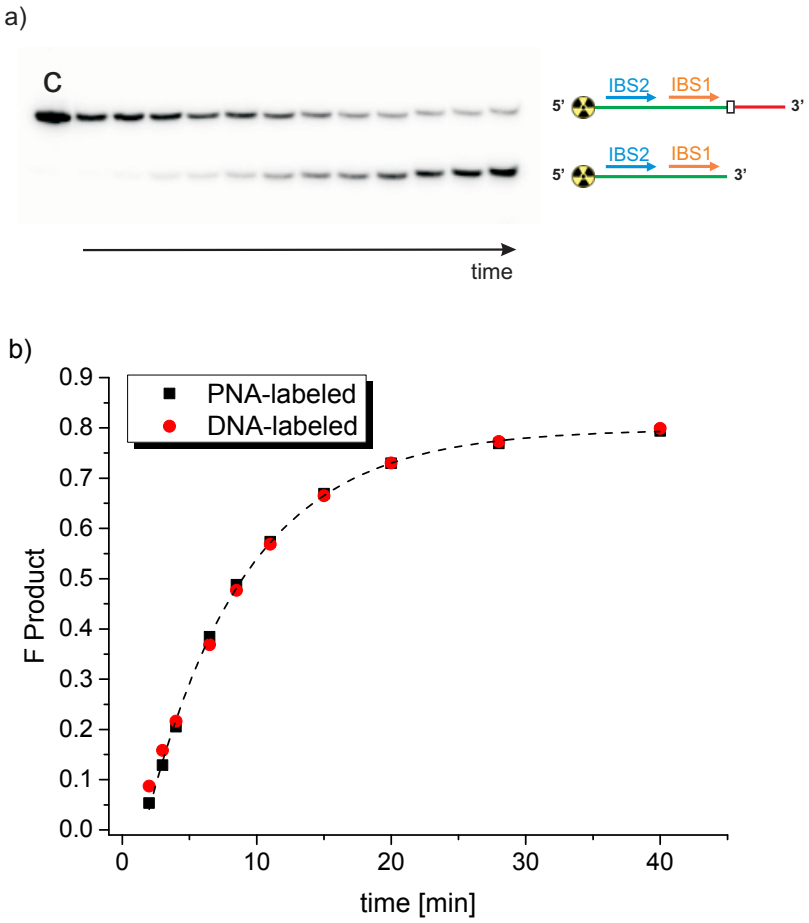


Figure 63: Cleavage of substrate 17/7 by D135-L14 labeled with DNA-dye or PNA-dye. Cleavage of [^{32}P]-5'-end labeled substrate 17/7 by D135-L14 with hybridized DNA-dye oligomers or PNA-dye oligomers. a) Cleavage bands of substrate at different reaction time points detected on denaturing PAGE. Control C is unreacted substrate. b) Plot of fraction of cleaved product versus the reaction time and exponential fit (dotted line). Comparison of experiments with DNA-labeled (red squares) and PNA-labeled (black dots) D135-L14.

As reported in the literature,[42] the modification of D135 in the two loops L1 and L4 for hybridization of DNA labels leads to a small decrease in activity (Table 5). D135-L14 with hybridized DNA-dye or PNA-dye oligonucleotides reveals a higher activity than D135-L14 without labels. The rate constants of D135-L1shL4 are similar to the ones of D135-L14. As described previously (Chapter 3.6.5) D135-L14sh contains a stable G-C rich stem replacing loop L4. This structural motif influences the folding behavior of the ribozyme and consequently a reduced activity is observed for D135-L14sh compared to the other constructs.

Table 5: Cleavage rate constants k_{obs} of D135, D135-L14, D135-L14sh and D135L1shL4 in min^{-1} determined from single exponential fits. The values given correspond to the arithmetic mean of at least three independent measurements with errors corresponding to one standard deviation.

RNA construct	Unlabeled	DNA labeled	PNA labeled
D135	0.109 ± 0.005	-	-
D135-L1shL4	0.100 ± 0.003	-	0.118 ± 0.002
D135-L14	0.098 ± 0.003	0.115 ± 0.003	0.118 ± 0.004
D135-L14sh	0.073 ± 0.002	-	0.088 ± 0.002

Besides the increased cleavage rate upon label hybridization also the fraction of cleaved substrate is increased by almost 10 % when DNA or PNA labels are hybridized to D135-L14 compared to unlabeled D135-L14 (Table 6).

Table 6: Maximum cleavage of substrate 17/7 by D135, D135-L14, D135-L14sh and D135-L1shL4 in %.

RNA construct	Unlabeled	DNA labeled	PNA labeled
D135	73	-	-
D135-L1shL4	71	-	76
D135-L14	76	84	85
D135-L14sh	69	-	74

Hence, the observed activity of RNA-PNA and RNA-DNA hybrids is similar, confirming that PNA does not reduce the activity or influences the folding behavior of the ribozymes, respectively.

3.6.8 Single-Molecule FRET Experiments

After demonstrating the efficient and non-destructive labeling by the use of PNA, the studies were completed by proving the applicability of the PNA labeling scheme in single-molecule FRET experiments. The group II intron model system has been investigated via smFRET using the D135-L14 construct with DNA-labels by Steiner *et al.* (Chapter 3.2.3).[42] Using the same conditions, the new labeling strategy with PNA labels was tested in comparison to DNA labels. Ideally, the new labeling scheme should not affect the folding pathway. The smFRET experiments were carried out with the D135-L14 and the new D135-L1shL4 model intron, prefolded at 42 °C in the presence of 100 mM Mg^{2+} (optimal folding conditions *in vitro*, Chapter 3.2.3), and labeled with the respective oligonucleotides (Table 7).

Table 7: DNA-Cy and PNA-Cy derivatives and label positions within the ribozyme RNA used for smFRET experiments.

Experiment	RNA	Label	Loop
DNA	D135-L14	**DNA1-Cy3**	L1
		DNA2-Cy5	L4
PNA3	D135-L14	**PNA4-Cy3**	L1
		PNA3-Cy5	L4
PNA3S	D135-L14	**PNA4-Cy3**	L1
		PNA3S-Cy5	L4
D135-L1shL4	D135-L1shL4	**PNA4-Cy3**	L1
		PNA3-Cy5	L4

The donor dye Sulfo-Cy3 was placed in loop L1 via hybridization of **DNA1-Cy3** or **PNA4-Cy3**, respectively, and the acceptor dye Cy5 at loop L4 *via* hybridization of **DNA2-Cy5**, **PNA3-Cy5** or **PNA3S-Cy5**. D135-L1shL4 was hybridized with **PNA4-Cy3**/**PNA3-Cy5** (Figure 58). The labeled RNA was then immobilized on a quartz surface via a biotin-streptavidin coupling (Chapter 1.2.3). For simplicity the experiments are named after their main difference (Table 7).

Characteristic time trajectories of each experiment with applied Hidden Markov modeling[174] (HMM) are depicted in Figure 64.

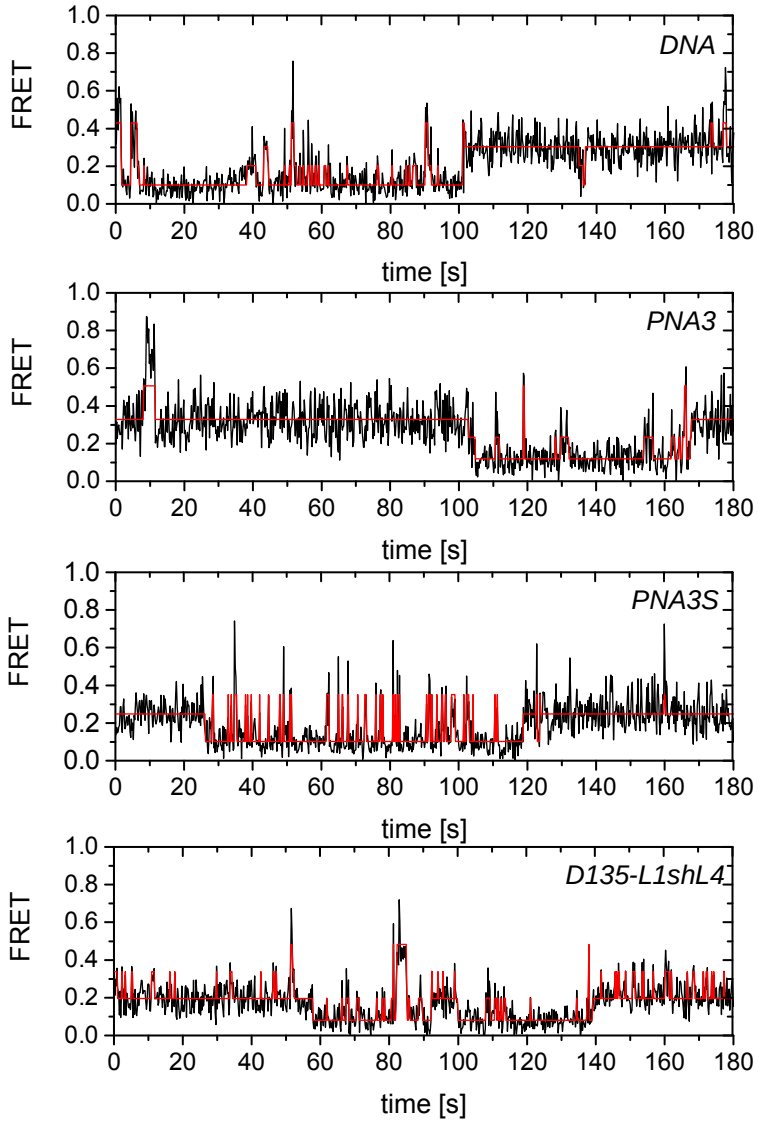


Figure 64: Characteristic smFRET time trajectories of experiments *DNA*, *PNA3*, *PNA3S*, and *D135-L1shL4*. The different conformational states at 0.1, 0.2, 0.4 and 0.6 are detected by the software method vbFRET based on HMM (red line).

For the DNA and the PNA dye labeling scheme, the same main FRET states of ≈ 0.1, 0.2, 0.4, and 0.6 were observed (Figure 64). Besides transitions with longer dwell times, which is the time a molecule spends in a state before transiting to another state, fast transitions are detected. Possible intermediate states of fast transitions cannot be resolved by the experimental setup and can therefore not be distinguished from direct transitions between low and high states, *e.g.* 0.1 to 0.4.

The observed states can be assigned to different RNA conformations from the unfolded state U via intermediates I_1, I_2 and F (0.1 0.2, 0.4) to a compact structure N (0.6) (Figure 65), known from the literature.[42] Differences to the literature values derive from applied filter corrections in the software. The 0.1 state could be assigned with the use of alternating laser

excitation (Chapter 5.4.7) and a refined background correction (20 darkest pixels per frame) as well as correction for donor bleed through and direct acceptor excitation.[42]

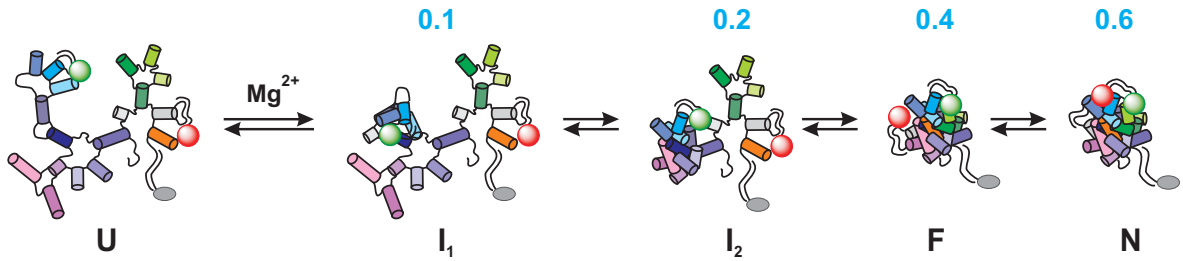


Figure 65: Suggested folding pathway of D135-L14. D135-L14 labeled with DNA or PNA oligomers carrying Sulfo-Cy3 (green), Sulfo-Cy3 (red), and biotin (grey). Compaction of the individual domains and corresponding FRET states (in blue) are shown. *Scheme is based on Ref. [42].*

The FRET histograms built from single-molecule time traces and analyzed by bootstrapping[175] revealed overall similar distributions of FRET states in experiments with DNA and PNA labeled D135-L14, and PNA labeled D135-L1shL4 (Figure 66).

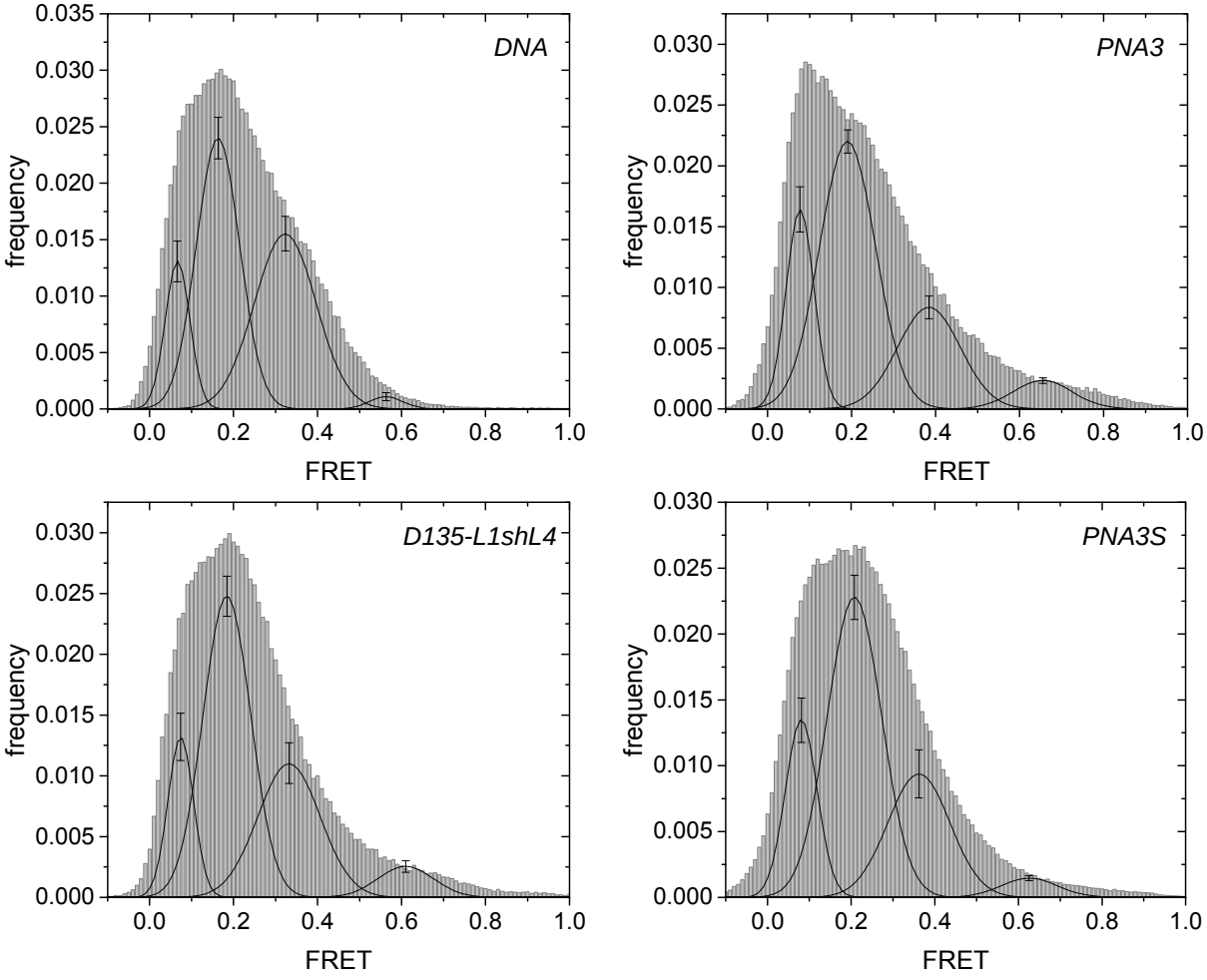


Figure 66: Normalized cumulated FRET histograms built from individual time trajectories of experiments *DNA*, *PNA3*, *PNA3S*, and *D135-L1shL4*. Histograms with bootstrapped Gaussian fits (black)[175] show the distribution of the different FRET states with standard deviation of the amplitude.

The relative occurrence of those FRET states (Figure 67 a) and their center of distribution determined by bootstrapped Gaussian fitting of the histograms (Figure 67 b), showed similar trends in all labeling schemes. However a slightly higher occurrence of the most compact state 0.6 can be observed together with a decrease of the 0.4 state in case for the PNA labeling compared to DNA labeling.

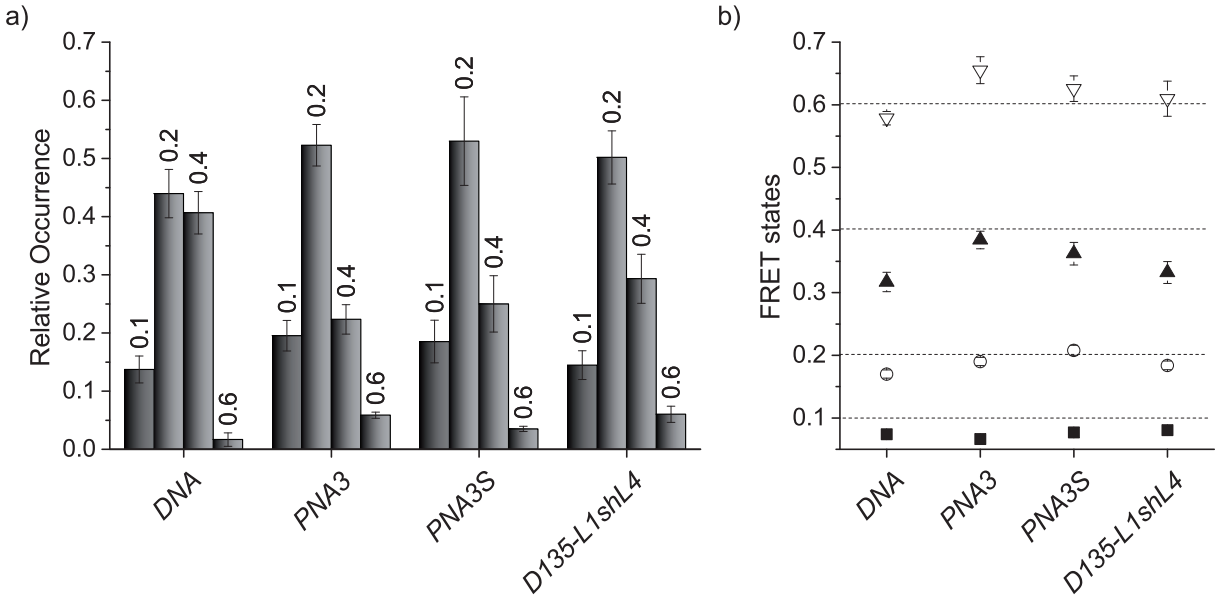


Figure 67: FRET states and their relative occurrence from smFRET experiments *DNA*, *PNA3*, *PNA3S*, and *D135-L1shL4*. a) Relative occurrence of FRET states 0.1, 0.2, 0.4, and 0.6 determined from bootstrapped Gaussian fits[175] of histograms, with standard deviation. b) FRET states given by center of distribution of each Gaussian fit with standard deviation. The dotted lines indicate the observed FRET states 0.1, 0.2, 0.4, and 0.6.

Besides the folding conformation of the RNA, its dynamic behavior was investigated. The occurrence of transitions between different states was calculated from the dwell times, and plotted as transition density plots (TDP, Figure 68).[43,174] For all labeling schemes, the main transitions were observed between the clustered states 0.1↔0.2, and 0.2↔0.4 (Figure 68). Additionally, in all experiments a second family of transitions (0.1↔0.2↔0.5) was observed. A mirror symmetry relative to the main diagonal shows the reversibility of the conformational changes. The folding of PNA labeled D135-L14 exhibited more dynamics observed in a broader distribution of the transition densities, especially in experiments with **PNA3-Cy5**. PNA-labeled D135-L1shL4 with shortened loop L1 was found to behave similarly to DNA-labeled D135-L14. The PNA derivative **PNA3S-Cy5** containing the spacer group between the PNA sequence and the Sulfo-Cy5 did not lead to a significantly changed behavior compared to the one without, **PNA3-Cy5**.

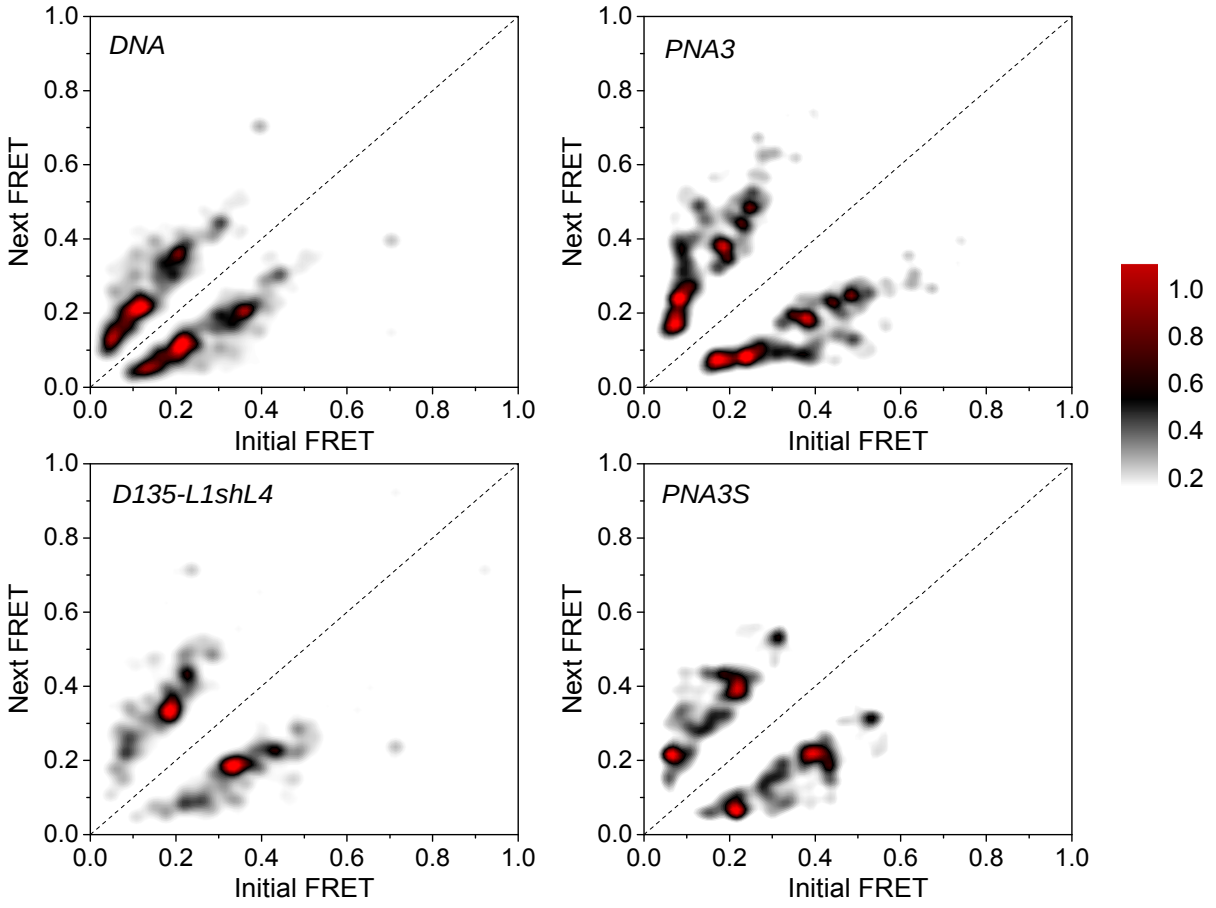


Figure 68: Transition density plots of smFRET experiments *DNA*, *PNA3*, *PNA3S*, and *D135-L1shL4*. TDPs calculated from the dwell-time analysis. The color bar depicts the normalized number of events. Main transitions are detected between 0.1 and 0.2, as well as 0.2 and 0.4.

Additional smFRET experiments were carried out using the acceptor PNA labels **PNA1S-Cy5** (***PNA1S***), **PNA2-Cy5** (***PNA2***) and **PNA2S-Cy5** (***PNA2S***), and the donor PNA label **PNA4-Cy3** hybridized to D135-L14. The derivatives carrying the acceptor were designed to hybridize to different positions within the loop L4, and differ in the presence of a spacer group (S) between the dye and the PNA. In the distribution histograms and TDPs, however, no significant differences compared to the other derivatives (**PNA3-Cy5** and **PNA3S-Cy5**) were observed (Figure 69). The only variation is that experiments with DNA labels or **PNA1S-Cy5** respectively show a distribution of the transition density closer to the diagonal. This indicates that the transitions occur mainly between FRET states close to each other, *e.g.* 0.1↔0.2 with small conformational changes. As seen before in experiments with **PNA3-Cy5** and **PNA3S-Cy5**, also no differences were observed with or without the presence of the space group.

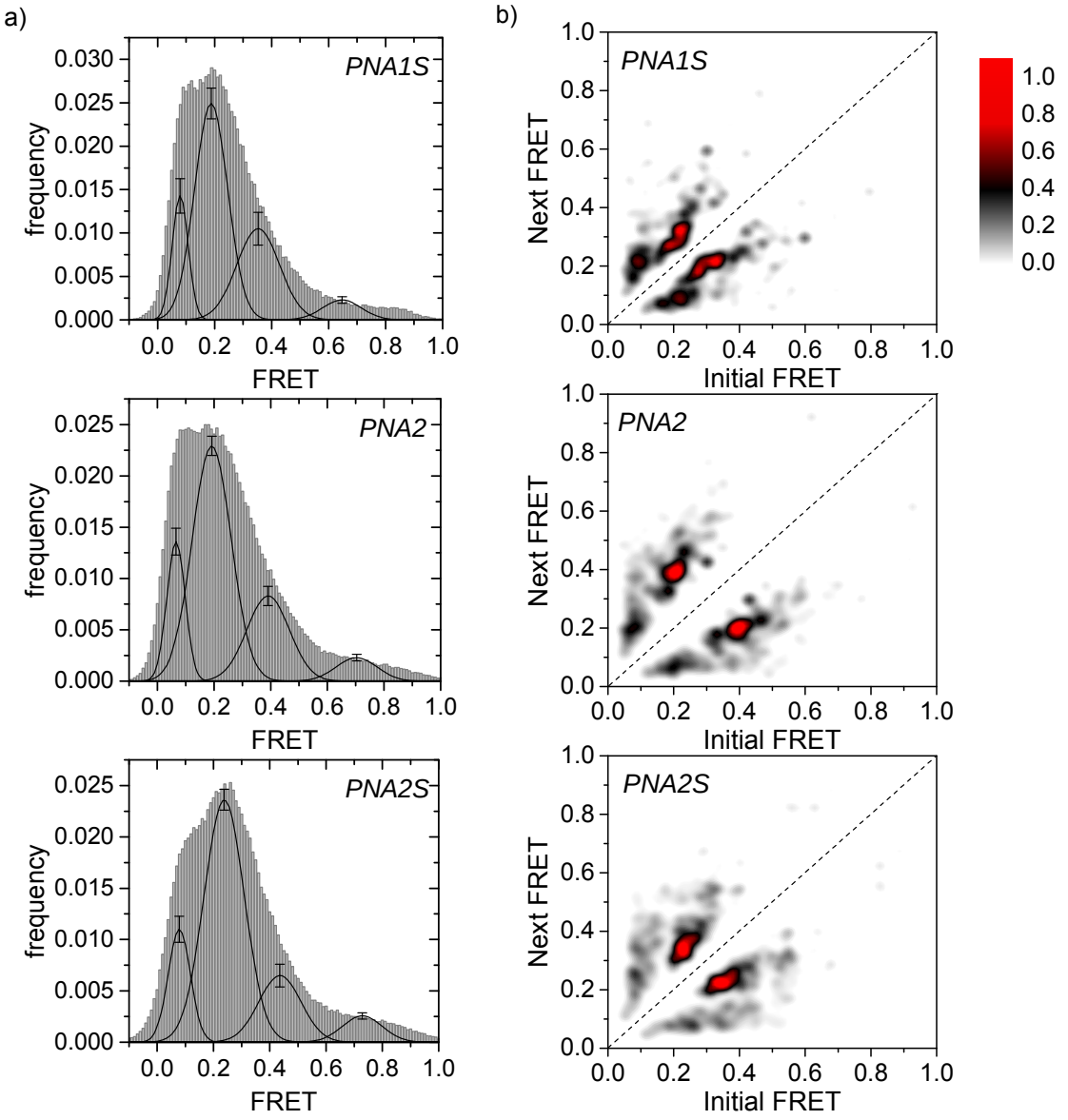


Figure 69: Normalized cumulated FRET histograms built from individual time trajectories and TDPs of smFRET experiments *PNA1S*, *PNA2,* and *PNA2S*. a) Histograms with bootstrapped Gaussian fits (black)[175] show the distribution of the different FRET states. b) TDPs calculated from the dwell-time analysis. The color bar depicts normalized number of events. Main transitions are detected between 0.1 and 0.2, as well as 0.2 and 0.4.

After investigations of the PNA labeling scheme and systematic comparison using a model system, the labeling scheme was applied to a new RNA system (following chapter).

3.6.9 Labeling of the *btuB* Riboswitch of *E. coli*

To investigate interactions of the *btuB* riboswitch with the novel fluorescent B_{12} derivatives (Chapter 2) by use of FRET, the RNA needs to be fluorescently labeled as well. The *btuB* riboswitch construct pPC4 by Choudhary *et al.*[176] was originally designed for labeling by the hybridization method introduced in Chapter 3.3. The construct comprises an artificial loop similar to L4 in D135-L14 which replaces the natural tetraloop in P11 (Figure 70, Appendix). An 18 nucleotide long DNA oligomer with a dye, *e.g.* Cy3 or Cy5 can thus be hybridized to the RNA. Fluorescence native gel studies were carried out to verify the labeling efficiency of a DNA-Cy5 oligonucleotide binding to the modified riboswitch pPC4. However, an insufficient labeling yield was detected by native fluorescence gels (Figure 70 b and c). In this way bulk FRET measurements are not possible without a further purification step, since unbound labeled DNA oligomers in solution lead to substantial background fluorescence.

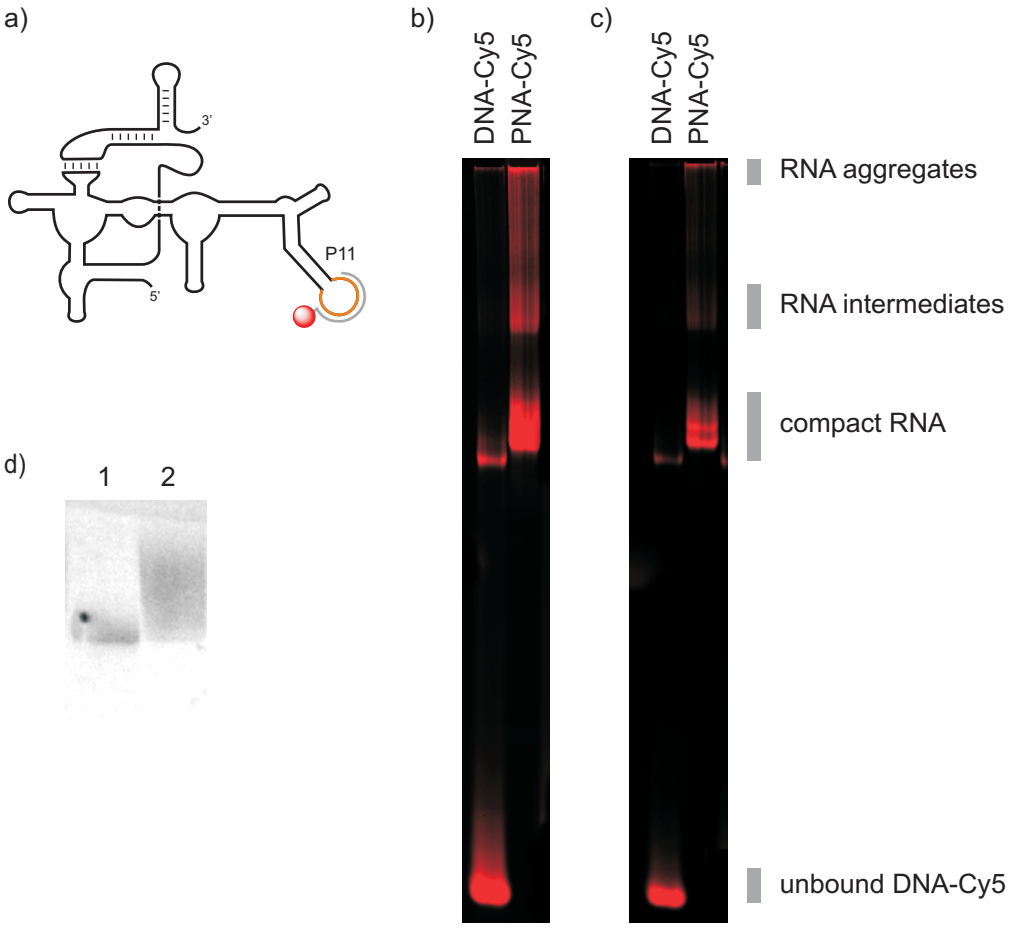


Figure 70: PAGE experiments of the *btuB* riboswitch construct pPC4. a) Schematic drawing of the pPC4 riboswitch RNA with a modification in loop P11 (orange) to hybridize oligonucleotide labels (grey) carrying Sulfo-Cy5 (red). Binding of DNA-Cy5 or PNA-Cy5 to pPC4 detected by fluorescence on native gels with b) high contrast and c) low contrast. d) Denaturing PAGE gel (10 %) of pPC4 RNA detected by UV shadowing. Lane 1: lower fraction of purification gel, lane 2: upper fraction of purification gel.

Furthermore, a misfolding of the RNA was detected in native and denaturing PAGE experiments. Denaturing PAGE experiments did not reveal distinct RNA bands but a broad smear. A refolding of the misfolded RNA by annealing for 60 s at 90 °C resulted in a more distinct band. However, in native PAGE experiments with the refolded RNA several RNA species were detected. A further purification of the RNA by denaturing PAGE excising only the lowest part of the band yielded one distinct band (Figure 70 d). The misfolding is most likely caused by the modified loop. The G-C rich stem leads to a very stable duplex and the loop itself is fairly large compared to other structural elements. Consequently, a correct folding of the riboswitch was not granted and the construct was not used for further studies. To overcome these drawbacks of low labeling yields and misfolding of the RNA, the investigated Peptide Nucleic Acid labeling strategy (*vide supra*) was used. This strategy gives the opportunity to drastically reduce the modifications of the RNA and thus avoid a misfolding, while at the same time increasing the labeling efficiency. Fluorescence native gel studies of **PNA1S-Cy5** hybridized to the modified loop in the pPC4 riboswitch RNA showed a significantly increased labeling efficiency compared to DNA-Cy5 (Figure 70 b and c).

3.7 Discussion and Conclusion

In Chapter 2 it was shown how a metabolite can be labeled with fluorophores to investigate its interaction with a functional RNA using the FRET technique. The present chapter, on the other hand, addressed how the RNA can be fluorescently labeled.

For fluorescence applications large RNAs are usually labeled indirectly by hybridization of fluorescent DNA oligonucleotides. Typically the oligonucleotides are ~18 nucleotides long to effectively label the RNA. Consequently, modifications within the RNA are necessary to provide a hybridization platform for the DNA oligomers. These modifications can lead to misfolding of the RNA. Furthermore, the labeling yields using this method can be low depending on the RNA system. Based on this the labeling strategy was improved by use of Peptide Nucleic Acids. New PNA derivatives carrying sulfonated Cy3 or Cy5 FRET dyes were synthesized by solid phase synthesis and further characterized. It is a quick and effective method to obtain individual sequences and allows introducing dyes specifically at the desired position. The PNA sequences were designed to label a model system, the group II intron ribozyme D135-L14, which has already been investigated with the DNA labeling scheme by smFRET. This allows a direct comparison of the two strategies.

Fluorescent probes are typically depended on environmental effects that can influence the fluorescence and consequently careful investigations of the photophysical properties are necessary. Fluorescence quantum yield and steady-state anisotropy experiments demonstrated that the main photophysical properties of the dyes do not change if the labeling system is varied from DNA oligomers to PNA oligomers. The effect of fluorescence enhancement of the cyanine dyes bound to nucleic acids is as well observed for PNAs. This is a consequence of the reduced photoisomerization process, which deactivates the first excited state. Furthermore steady-state anisotropy measurements reveal a similar behavior as observed for the attachment of Cy3 and Cy5 to B_{12} (Chapter 2). The anisotropy value of Cy3 is unchanged whereas the value upon Cy5 attachment is increased compared to the free dyes. This indicates once more that steady-state anisotropy can be rather insensitive and time-resolved studies such as lifetime measurements and dynamic anisotropy could give more details.[20,21,28]

In melting experiments it was shown that short PNAs with a length of 10 nts form a more stable duplex with RNA than the 18 nts long DNA oligonucleotides. This suggests that even shorter PNA sequences are possible. However, system depended sequence specificity has to be considered for each RNA as shorter sequences are more likely to be repeated in the RNA sequence. The shorter PNA hybridization probes allow minimizing the modifications within the RNA necessary for the binding of the label. To underline this, the artificially elongated loop in the intron model system D135-L14, to which the labels are bound, was

shortened to make the construct more similar to the natural structure, yielding the new construct D135-L1shL4. Additional shortening of loop L4 led to a decreased ribozyme cleavage activity and a low labeling efficiency due to base pair formation in the loop leading to an unfavorable structure. This can be avoided by a redesign of the loop with a non self-complementary single-stranded structure.

DNA oligonucleotides typically need to be annealed to the RNA at high temperatures (90 °C) at which the RNA is fully denatured. Due to the stronger binding affinity of PNA the hybridization procedure could be optimized. The PNA is hybridized after the folding of the ribozyme at 42 °C. Consequently, dye decomposition and RNA degradation can be reduced as well as RNA aggregation, which may occur while annealing at high temperatures. In fluorescence native gels a higher labeling efficiency of PNA derivatives over DNA was observed in a higher fraction of labeled RNA and less unbound label. No excess of PNA is needed compared to DNA. With the high labeling efficiencies of the PNAs further purification steps from the unbound label, which decrease the labeling yield, can be avoided compared to DNA. Consequently the superior labeling properties of PNA reduce background issues in fluorescence applications and give a higher fraction of detectable molecules with less amount of probe needed.

In cleavage assays mimicking the first step of splicing was confirmed that folding and activity of the ribozyme RNA remain basically unaffected by the PNA label, and are similar to the labeling with DNA.

In smFRET experiments the same FRET states 0.1, 0.2, 0.4 and 0.6 and overall similar trends of the state distributions were observed for DNA or PNA labeled ribozyme. Marginal variations in the FRET states could derive from different spatial positions of the fluorophores. Moreover, Gaussian fitting of such a complex system with overlapping distributions leads to fairly large deviations. The lower occurrence of the 0.4 state and higher occurrence of the 0.6 state in experiments with PNA compared to DNA shows that the RNA spends less time in this intermediate state. Consequently, with the shorter PNA sequences and a shortened loop L1 the RNA favors the more compact structure. Nevertheless, the folding behavior of the RNA is not changed by the PNA. Also, the dynamics of the RNA system show similar trends in all labeling schemes. In TDPs, main transitions between the states 0.1↔0.2↔0.4 were observed. The transitions to higher FRET states, *i.e.* 0.6, were detected less often due to a lower occurrence of the high state. Also slower kinetics are generally underrepresented in TDPs.[43] Additionally in all experiments a second family of transitions (0.1↔0.2↔0.5) is observed which might derive from an alternative folding pathway. The RNA folding of PNA labeled D135-L14 exhibited slightly more dynamics observed in a broader distribution of the transition densities. This might also be due to the more flexible loop structure. The loop was designed to bind 18 nts, but the PNA sequences are drastically shorter (10 nts). With the

shortened loops adjusted to the shorter sequences the dynamics are decreased and again similar to DNA labeling. Underlining this hypothesis the density distribution around the diagonal while labeled with the 14-mer PNA derivative is similar to DNA labeling. Hence, as observed for the folding, also variations in the dynamics do not derive from the DNA or PNA, respectively. More likely small variations are due to differences in the loop structure where the labels are bound, and from the general folding heterogeneity of the construct. The latter can also be observed in the native fluorescence gel studies by broadening of the bands. It is not surprising that complex systems like this large group II intron ribozyme show a high variability, *e.g.* due to a sample conformational heterogeneity. With the presented studies could be demonstrated that PNA labels can be used for smFRET experiments since they do not change the folding and dynamics of the RNA in comparison to DNA.

To evaluate the possible influence of the distance between the fluorophores and the PNA the experiments were carried out with different derivatives with or without a spacer group between the cyanine dye and the PNA. No significant differences in the duplex stability, rotational freedom, fluorescence emission or folding of the RNA were observed.

Finally, the investigated PNA labeling strategy was applied to the *btuB* riboswitch of *E.coli*. The labeling efficiency was remarkably higher using a labeling with PNA instead of DNA. This system suffers from large modifications which can consequently be reduced or avoided with the PNA labeling scheme.

In this proof of concept study, we demonstrated that PNA oligonucleotides can be used as specific internal labeling agents for large RNA systems to study their folding mechanism without influencing the latter. Consequently, PNA-dye derivatives can be employed as a fast, efficient and easy to handle labeling probes for various fluorescence applications including smFRET for imaging of processes *in vitr*o. Moreover, as RNA-PNA hybrids are not affected by RNase H degradation, contrary to RNA-DNA, they are suitable for RNA folding studies *in vivo*.[155]

In a next step, one could envision systems that are labeled at natural occurring positions without further modifications. At the same time, we observed that the folding and dynamics depend on the structure of the modification where the label is bound. These findings are important for future studies of various nucleic acid systems to avoid structural modifications of natural RNA and hence hybridize PNA directly to existing structural elements. PNA labeling has a great potential in particular for systems, where modifications play a tremendous role, *e.g.* the *btuB* riboswitch of *E. coli*.

4 Collaborations

4.1 Imaging of an RNA-Splicing Process Using Fluorescent PNA Labels

The following chapter is based on collaboration with Susann Zelger-Paulus in the group of Prof. Roland K. O. Sigel at the University of Zürich.

4.1.1 Introduction

Herein, an application of the investigated labeling strategy for RNA by using PNA hybridization probes described in Chapter 3 is shown. The RNA system of interest is the group II intron *Sc.*ai5γ originated from *Saccharomyces cerevisiae*. As described previously this RNA folds in a defined three dimensional structure and actively induces its self-cleavage reaction during RNA maturation. PNA labels are used in these studies to investigate the splicing process by bulk FRET measurements, fluorescence microscopy and native fluorescent gel assays. To this end, PNAs carrying the relevant FRET dyes, either Sulfo-Cy3 or Sulfo-Cy5, are hybridized to the intron flanking regions. FRET occurs when the intron is self-spliced as the exons are ligated and the two fluorophores attached to the PNA come nearby (Figure 71). With regard towards *in vivo* studies of the splicing mechanism, this labeling strategy is very promising.

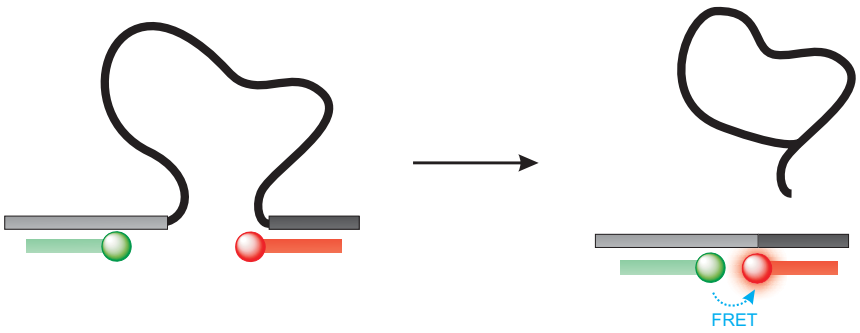
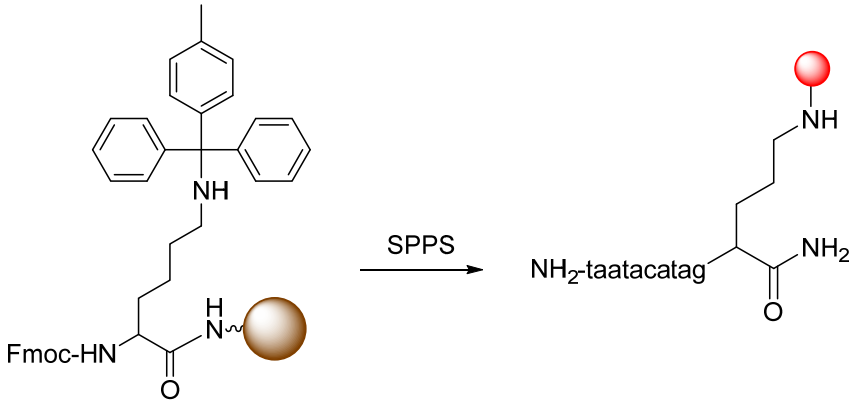


Figure 71: Imaging of the splicing process of a group II intron ribozyme with fluorescent PNA labels. The exon sites (grey bars) of a ribozyme (black) are labeled by hybridization of fluorescent PNA oligomers (green donor, red acceptor). If splicing occurs donor and acceptor get close and FRET occurs.

4.1.2 Synthesis and Characterization of Fluorescent PNA Derivatives

Two PNA derivatives **PNA-Ex1-Cy3** and **PNA-Ex2-Cy5** were designed to carry the fluorescent cyanine dyes Sulfo-Cy3 or Sulfo-Cy5 (FRET pair), and to be complementary for hybridization to the intron flanking regions of the group II intron *Sc*.ai5γ.

The PNA derivatives were prepared by use of solid phase synthesis (see Chapter 3).[170] The cyanine dye Sulfo-Cy3 was covalently attached at the *N*-terminus of the PNA sequences **PNA-Ex1**, and Sulfo-Cy5 at the *C*-terminus of the PNA sequences **PNA-Ex2**. Additionally lysine was used to increase the solubility. For *in vivo* studies a signal peptide for the delivery into mitochondria can be attached to the PNA. To keep the *N*-terminus of **PNA-Ex2-Cy5** available for the peptide coupling and have the dye at the right position for the necessary FRET distance, the synthetic procedure was adjusted. The resin was functionalized with *N*-α-Fmoc-*N*-ε-4-methyltrityl-L-lysine (Fmoc-Lys(Mtt)-OH) to allow coupling of Sulfo-Cy5 to the *C*-terminus of the PNA sequence. The *N*-ε-Mtt function was removed under conditions (1 % TFA in DCM) at which tert-butyl type groups and peptide ester bonds to the TFA-labile resin remain intact.[177]



Fmoc-Lys^{Mtt}-resin **PNA-Ex2-Cy5**

Scheme 8: Synthesis of PNA-Ex2-Cy5. Starting from a Lys^{Mtt} functionalized resin (brown) **PNA-Ex2-Cy5** (Sulfo-Cy5 red) was synthesized by SPPS.

PNA-Ex1-Cy3 and **PNA-Ex2-Cy5** derivatives were characterized by HPLC, ESI-MS and MALDI-MS. Chromatograms of **PNA-Ex1-Cy3** reveal peaks at 260 nm (absorbance PNA), 550 nm (absorbance Sulfo-Cy3) and emission at 570 nm (excitation at 550 nm) with a retention time of 9.0 min. **PNA-Ex2-Cy5** is detected at 260 nm, 650 nm (absorbance Sulfo-Cy5), and emission at 670 nm (excitation at 650 nm) with a retention time of 8.8 min (Figure 72).

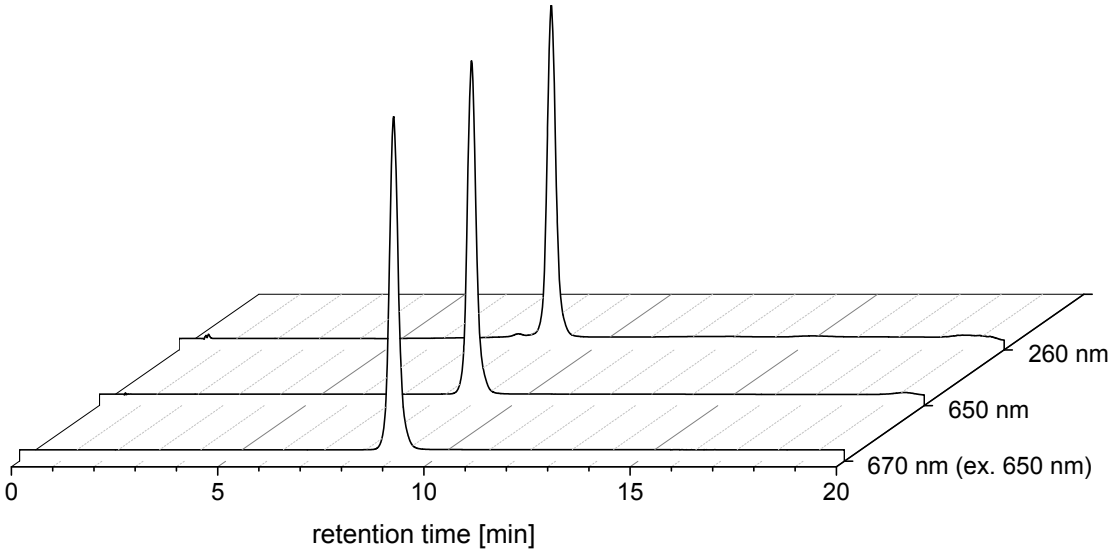


Figure 72: HPLC chromatograms of PNA-Ex2-Cy5 at different wavelength. At 260 nm the absorption of the PNA is detected, at 650 nm the absorption of the Sulfo-Cy5. The Sulfo-Cy5 emission is detected at 670 nm upon excitation at 650 nm. The intensities are normalized.

In MALDI-MS measurements **PNA-Ex1-Cy3** is detected at m/z = 3492.4 (calcd. for $C_{145}H_{185}N_{64}O_{38}S_2^+$ ($[M+2H]^+$) 3492.4) and **PNA-Ex2-Cy5** at 3486.5 (calcd. for $C_{145}H_{183}N_{64}O_{36}S_2^+$ ($[M+2H]^+$) 3486.4). In ESI-MS are the characteristic differently charged species observed due to multiple protonations of the PNA oligomers.

4.2 RNA Interaction with Rhenium Compounds

The following chapter is based on the publication:

"Photo-induced uncaging of a specific Re(I) organometallic complex in living cells"

by Anna Leonidova, Vanessa Pierroz, Riccardo Rubbiani, Yanjun Lan, Anita G. Schmitz, Andres Kaech, Roland K. O. Sigel, Stefano Ferrari, and Gilles Gasser in **Chemical Science**, 2014.[178]

4.2.1 Introduction

In the last decades, a large number of organometallic complexes have shown promising anti-proliferative activity towards different cancer cell lines. However, these compounds generally had low cellular uptake and low selectivity towards cancer cells over healthy cells. The use of external triggers (*e.g.* light, ultra-sound, temperature, *etc.*) to modify the cytotoxic effect of a prodrug and the coupling of a targeting vector (*e.g.* peptides, antibodies, *etc.*) to a drug were found to be very successful techniques to tackle these drawbacks. Here, we envisioned combining these two methods, namely an external trigger (*i.e.* light activation) and a targeting vector, in an organometallic compound.

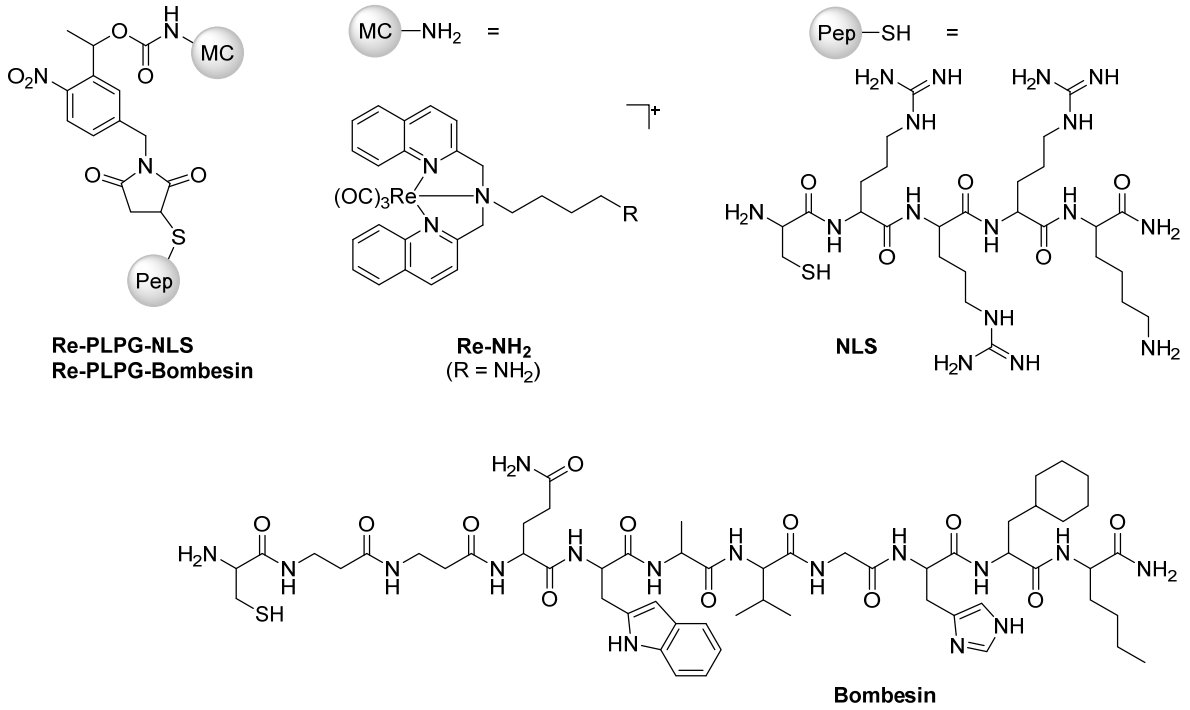


Figure 73: Structure of bioconjugated organometallic photo-caged derivatives Re-PLPG-NLS and Re-PLPG-Bombesin.

More specifically, a Re(I) tricarbonyl *N*,*N*-bis(quinolinoyl) complex (**Re-NH$_2$**) was derivatised with a photo-labile protecting group (PLPG) to cage **Re-NH$_2$** by formation of **Re-PLPG**. For organelle/cellular specificity, **Re-PLPG** was then further coupled to a nuclear localization sequence (NLS) or a bombesin peptide derivative to give **Re-PLPG-NLS** or **Re-PLPG-Bombesin**, respectively (Figure 73).

Photolysis experiments in *PBS buffer* (pH 7.4) demonstrated that **Re-NH$_2$** was completely photo-released from **Re-PLPG-NLS** and **Re-PLPG-Bombesin** using a very low irradiation dose (1.2 J cm^{-2}). To the best of our knowledge, these are the first two examples of the selective photo-release of an intact organometallic compound from a bioconjugate. Of high interest, both derivatives showed toxicity comparable to that of cisplatin towards cervical cancer cells (HeLa) upon light irradiation, although the phototoxic index (PTI) varied greatly with the targeting peptide. The cell death mechanism of **Re-PLPG-NLS** was explored using different techniques, including fluorescence microscopy, ICP-MS, gel electrophoresis, flow cytometry and transmission electron microscopy (TEM). It could be demonstrated that HeLa cells treated with **Re-PLPG-NLS** in the dark and upon irradiation showed severe cell stress (nucleolar segregation, pyknosis and vacuolation). The data obtained from an Annexin V/propidium iodide (PI) assay indicated that, after an early apoptotic stage, the onset induced by **Re-PLPG-NLS** led to cell death, with features ascribable to late apoptosis and necrosis, which were more marked for the treatment involving irradiation.

4.2.2 Effect of Re-PLPG-NLS on Nucleic Acids

As RNA is one of the major components of nucleoli, the effect of **Re-PLPG-NLS** on a 638 nucleotide long RNA sequence (D135-L14 ribozyme derived from the *Sc*.ai5γ group II intron of *S. cerevisiae*, see Chapter 3)[42] was investigated. Denaturing polyacrylamide gel electrophoresis revealed a considerable decrease in intensity and shift of the RNA band due to **Re-PLPG-NLS**/RNA interactions and formation of non-migrating RNA agglomerates (Figure 74) at high concentrations of **Re-PLPG-NLS** (100 µM). This effect occurred both in the absence and the presence of light, and no difference was observed for UVA-irradiated samples.

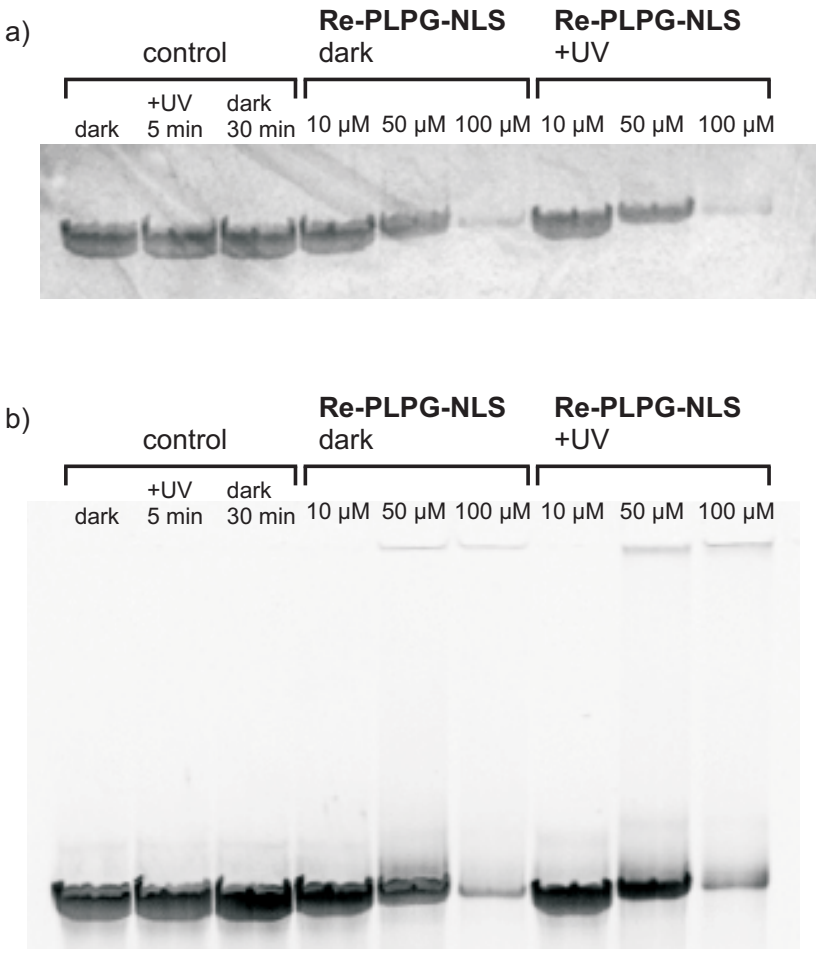


Figure 74: Denaturing polyacrylamide gel electrophoresis of D135-L14 ribozyme from *S. cerevisiae* incubated with Re-PLPG-NLS. Denaturing polyacrylamide gel electrophoresis of D135-L14 incubated with a) **Re-PLPG-NLS** visualized by UV shadowing, b) **Re-PLPG-NLS** visualized by ethidium bromide staining of RNA bands detected by UV shadowing at 260 nm.

Re-NH_2 or **Re-PLPG** alone did not affect the RNA (Figure 75). Given the localization of **Re-PLPG-NLS** in RNA-rich nucleoli, it is possible that the strong dark toxicity of this compound is linked to its effect on RNA.

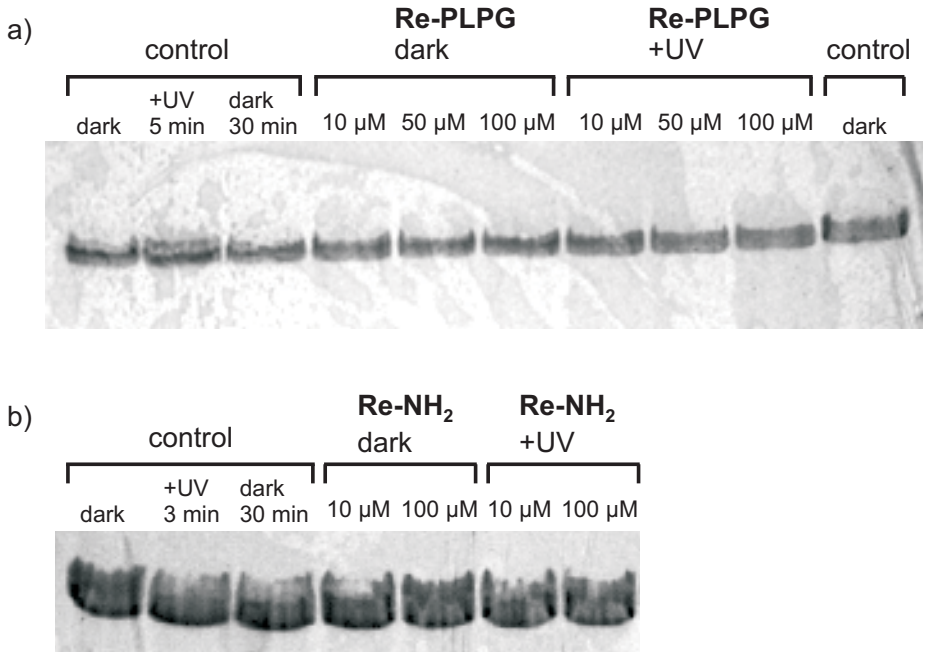


Figure 75: Denaturing polyacrylamide gel electrophoresis of D135-L14 ribozyme from *S. cerevisiae* incubated with Re-PLPG or Re-NH_2. Denaturing polyacrylamide gel electrophoresis of D135-L14 incubated with a) **Re-PLPG**, and b) **Re-NH_2**, visualized by UV shadowing.

5 Experimental Section

5.1 Instrumentation and Materials

All chemicals and solvents were of reagent-grade purity or better, and used as obtained from the commercial suppliers Sigma-Aldrich, and Thermo Fisher Scientific. Peptide nucleic acid (PNA) monomers and [2-(2-(Fmoc-amino)ethoxy)ethoxy]acetic acid (Fmoc-AEEA-OH spacer) were purchased from Link Technologies Ltd. (Lanarkshire, Scotland), and TentaGel S RAM-Lys(Boc)Fmoc resin beads from Rapp Polymers GmbH (Tübingen, Germany). (Balgach, Switzerland). Cy3 N-hydroxysuccinimide (NHS) ester, Cy5 NHS ester, sulfonated (Sulfo) Cy3 NHS ester, and Sulfo-Cy5 NHS ester were purchased from Lumiprobe (Hallandale Beach, USA).

Thin layer chromatography (TLC) was performed with Merck TLC Silicagel 60 F_{254} plates and visualized by UV light (254, 365 nm). Flash chromatography was carried out using Merck silica gel 60 (0.04-0.063 mm).

Enzymes and purification kits were obtained from Promega, Thermo Scientific or Qiagen. DNA-oligonucleotides with either Sulfo-Cy3 (5'-Sulfo-Cy3-ACCAAGAGCGTTATTAAT-3' ≡ **DNA1-Cy3**), Sulfo-Cy5 (5'-Sulfo-Cy5-ACGTAGTCCGAAATATAT-3' ≡ **DNA2-Cy5**) or Biotin (5'-Biotin-TGCATGCCTGCAGGTCGACTCTA-3' ≡ DNA-biotin) were purchased from IBA (Göttingen, Germany). RNA oligonucleotides (HPLC grade) for UV/Vis melting experiments and primer for PCR were obtained from Microsynth (Balgach, Switzerland).

Water was treated with a TKA GenPure water purification system and autoclaved. All buffers and salt solutions were filtrated through 0.2 µm filters before use. RNA isolation was performed using an Elutrap electro-elution apparatus with BT1 and BT2 membranes from Whatman (London, UK). Accugel (40 % (w/v), 29:1 Acrylamide-Bisacrylamide) and 10xTBE (0.89 M Tris Borate pH 8.3, 0.2 M Na_2EDTA) buffer from National Diagnostics (Atlanta, USA) were used for polyacrylamide gel electrophoresis (PAGE). Gels were visualized with a Bio-Vision system from Vilber (Eberhardzell, Germany). For radioactive labeling of the RNA [γ-^{32}P]-ATP (150 mCi/mL, 6000 Ci/mmol) from PerkinElmer (Schwerzenbach, Switzerland) was used. Scintillation counts were performed on a scintillation counter 22000CA Liquid Scintillation Analyzer from Canberra Packard. The Centrifuges Sorvall RC 6 Plus from Thermo Scientific with rotors SA-600, SH-3000 and SLA-3000, and Eppendorf 5804R with rotor A-4-44 were used. Samples were vacuum dried with a Concentrator 5301 from Eppendorf or a freeze dryer Alpha 2-4 LDplus from Christ (Osterode am Harz, Germany).

pH-Measurements were carried out using a Metrohm (Herisau, Switzerland) 605 pH-meter and a Minitrode glass electrode from Hamilton.

Sample irradiation in the UV-reactor was done with a Rayomet RMR-200 photo-reactor.

HPLC analyses were performed on a Dionex Ultimate 3000 system (Olten, Switzerland; now part of Thermo Fisher Scientific) equipped with a DAD and fluorescence detector, or on a Hitachi LaChrom Elite system. Analytical HPLC was done with a Waters XTerra RP8 5 µm column. Preparative HPLC was carried out on a Varian ProStar sytem with two solvent delivery system units model 215 and a UV-Vis detector model 320 equipped with a Zorbax Prep HT 300SB-C18 (5 µm) column from Agilent or a Waters XTerra Prep RP8 (5 µm) column.

MALDI-MS spectra were recorded on a Bruker Autoflex I, ESI-MS and LC-MS spectra on a Bruker Daltonics HCT instrument in combination with an Acquity Ultra Performance LC system from Waters equipped with a Zorbax 300SB-C18 column (3.5 µm) from Agilent.

NMR spectra were recorded on a Bruker Avance 500 spectrometer using a cryo BBO probe and referenced to an internal standard of residual deuterated solvent.

UV/Vis measurements were carried out on Varian, now part of Agilent (Basel, Switzerland), UV Cary 100 Bio and UV Cary 500 Scan instruments or a PerkinElmer Lamda 25 spectrophotometer using a quartz cuvette of 10 mm path length. Concentrations of nucleic acids were determined using a Nanodrop 2000 from Thermo Scientific.

Fluorescence spectra were recorded on a Varian Cary Eclipse fluorescence spectrophotometer, equipped with an automated polarizer.

Fluorescence native gels and storage phosphor screens were scanned on a Typhoon Scanner FLA-9500 and quantified with Image Quant TL from GE Healthcare.

Single-molecule FRET measurements were performed on a total internal reflection fluorescence microscope (TIRFM) system using a setup according to Rueda and coworkers.[41] The setup mainly includes an IX71 microscope from Olympus, a green laser at 532 nm and a red laser 640 nm from CrystaLaser (Reno, USA). Signals were recorded with an EMCCD camera DU 897 from Andor (South Windsor, USA). Data analysis was done using the home-written MATLAB-based software MASH (Multifunctional Analysis Software for Heterogeneous smFRET data).[179]

5.2 Buffers

The pH was adjusted by addition of either NaOH or HCl.

Reaction Buffer: 80 mM MOPS (pH 6.9), 500 mM KCl.

Elution Buffer: 10 mM MOPS (pH 6.0), 1 mM EDTA, 250 mM NaCl.

Transcription Buffer: 40 mM Tris-HCl (pH 7.5), 40 mM DTT, 2 mM spermidine, 0.01 % Triton X-100.

ME Buffer: 10 mM MOPS (pH 6.0), 1 µM EDTA.

Denaturing Loading Buffer: 11.7 M urea, 40 mM Tris-HCl (pH 7.5), 0.1 % xylene cyanol (XC), 0.1 % bromphenol blue (BB), 230 mM sucrose, 0.8 mM EDTA pH 8.0.

Native Buffer: 3 mM MgOAc, 66 mM HEPES, 34 mM Tris-HCl (pH 7.4).

PBS Buffer: 0.1 M phosphate buffer (pH 7.4), 0.1 M NaCl.

5.3 Methods Chapter 2 (Labeling of B_{12})

5.3.1 General Methods

HPLC method 1a (analytical): column Waters XTerra RP8; flow-rate 1 mL/min; detection wavelength 360 nm, 550 nm or 650 nm, excitation at 550 nm and emission at 570 nm or excitation at 650 nm and emission at 670 nm; solvent A: Triethylammonium acetate TEAA (100 mM, pH 7), solvent B: ACN/ H_2O 9:1; gradient: 0-2 min 0 % B, 2-20 min 0-55 % B, 20-23 min 55-0 % B, 23-25 min 0 % B.

HPLC method 1b (analytical): column Waters XTerra RP8; flow-rate 1 mL/min; detection wavelength 360 nm or 600 nm; solvent A: Triethylammonium acetate TEAA (100 mM, pH 7), solvent B: ACN/ H_2O 9:1; gradient: 0-2 min 0 % B, 2-20 min 0-55 % B, 20-25 min 55-5 % B, 25-27 min 5-0 % B, 27-28 min 0 % B.

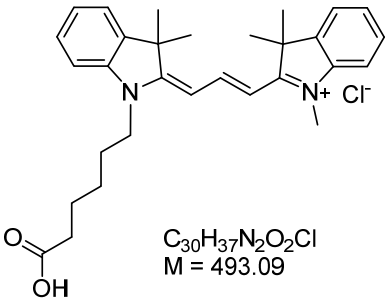
HPLC method 2 (preparative): column Waters XTerra RP8; flow-rate 30 mL/min; detection wavelength 360 nm; solvent A: Triethylammonium acetate TEAA (100 mM, pH 7), solvent B: ACN/ H_2O 9:1; gradient: 0-5 min 0 % B, 5-30 min 0-50 % B, 30-35 min 50-5 % B, 35-38 min 5-0 % B, 38-40 min 0 % B.

Purities were determined from the HPLC UV/Vis traces at 360 nm, 550 nm or 650 nm.

The atom numbering used for NMR peak assignment of B_{12} is shown in Figure 15. Reactions and analyses of B_{12} derivatives and dyes were done in the dark to prevent from decomposition or deactivation.

5.3.2 Synthesis of B_{12} Derivatives

Cy3: 2-[3-[1-(5-Carboxypentyl)-1,3-dihydro-3,3-dimethyl-2*H*-indol-2-ylidene]-1-propenyl]-1,3,3-trimethyl-3*H*-indolium



$C_{30}H_{37}N_2O_2Cl$
M = 493.09

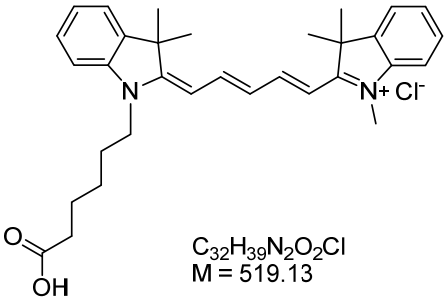
2-Methylene-1,3,3-trimethylindoline (0.12 mL, 0.65 mmol), *N,N'*-diphenylformamidine (0.13 g, 0.65 mmol), 1-(5-carboxypentyl)-2,3,3-trimethyl-3*H*-indolium chloride (0.23 g, 0.65 mmol) and sodium acetate (67 mg, 0.81 mmol) were dissolved in ethanol (50 mL). The reaction mixture was stirred at 85 °C for 24 h. The solvent was removed under reduced pressure and the crude reaction mixture was purified by flash chromatography (silica gel, DCM/MeOH) to give **Cy3** (55 mg, yield: 17 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 8.54 (t, $J_{a,b}$ = $J_{b,c}$ = 13.5 Hz, 1 H, **b**), 7.54 (d, J = 7.4 Hz, 2 H, **4**, **4'**), 7.44 (m, 2 H, **6**, **6'**), 7.36 (m, 2 H, **7**, **7'**), 7.31 (m, 2 H, **5**, **5'**), 6.48, 6.45 (dd, $J_{a,b}$ = $J_{b,c}$ = 13.5 Hz, 2 H, **a**, **c**), 4.15 (t, J = 7.5 Hz, 2 H, **NC*H*$_2$**), 3.69 (s, 3 H, **NC*H*$_3$**), 2.24 (t, J = 7.3 Hz, 2 H, **COC*H*$_2$**), 1.86 (m, 2 H, **NCH$_2$C*H*$_2$**), 1.76 (s, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**), 1.70 (m, 2 H, **COCH$_2$C*H*$_2$**), 1.52 (m, 2 H, **NCH$_2$CH$_2$C*H*$_2$**) ppm.

^{13}C{^{1}H}-NMR (125 MHz, MeOD-d_4, 300 K): δ = 180.58 (***C*O**), 176.55, 175.93 (**2**, **2'**), 152.06 (**b**), 144.09, 143.36 (**7a**, **7a'**), 142.17, 142.06 (**3a**, **3a'**), 130.00, 129.93 (**6**, **6'**), 126.70 (2 C, **5**, **5'**), 123.49, 123.39 (**4**, **4'**), 112.48, 112.24 (**7**, **7'**), 103.82, 103.71 (**a**, **c**), 50.62, 50.54 (**3**, **3'**), 45.13 (**NC*H*$_2$**), 37.33 (**COC*H*$_2$**), 31.76 (**NC*H*$_3$**), 28.29, 28.17 (**3-C*H*$_3$, 3'-C*H*$_3$**), 27.56 (**NCH$_2$C*H*$_2$**), 27.11 (**NCH$_2$CH$_2$C*H*$_2$**), 26.66 (**COCH$_2$C*H*$_2$**) ppm.

ESI-MS m/z: 457.3 $[M]^+$.

Cy5: 2-[5-[1-(5-Carboxypentyl)-1,3-dihydro-3,3-dimethyl-2*H*-indol-2-ylidene]-1,3-pentadien-yl]-1,3,3-trimethyl-3*H*-indolium



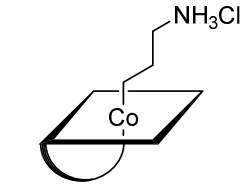
$C_{32}H_{39}N_2O_2Cl$
M = 519.13

2-Methylene-1,3,3-trimethylindoline (0.95 mL, 5.36 mmol), *N*-((1*E*)-3-(phenylimino)prop-1-enyl)benzenamine hydro-chloride (1.38 g, 5.36 mmol), 1-(5-carboxypentyl)-2,3,3-trimethyl-3*H*-indolium chloride (1.90 g, 5.36 mmol) and sodium acetate (0.53 g, 6.51 mmol) were dissolved in ethanol (300 mL). The reaction mixture was stirred at 85 °C for 24 h. The solvent was removed under reduced pressure and the crude reaction mixture was purified by flash chromatography (silica gel, DCM/MeOH) to give **Cy5** (yield could not be determined).

^{1}H-NMR (500 MHz, D_2O, 300 K): δ = 7.90 (t, $J_{a,b} = J_{b,c} = J_{c,d} = J_{d,e}$ = 12.6 Hz, 2 H, **b**, **d**), 7.41 (m, 2 H, **4**, **4'**), 7.32 (m, 2 H, **6**, **6'**), 7.17 (m, 4 H, **7**, **7'**, **5**, **5'**), 6.45 (t, $J_{b,c} = J_{c,d}$ = 11.3 Hz, 1 H, **c**), 6.14 (m, *2* H, **a**, **e**), 3.96 (m, 2 H, **NC*H*$_2$**), 3.46 (s, 3 H, **NC*H*$_3$**), 2.06 (m, 2 H, **COC*H*$_2$**), 1.72 (m, 2 H, **NCH$_2$C*H*$_2$**), 1.56 (s, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**), 1.30 (m, 2 H, **COCH$_2$C*H*$_2$**), 1.16 (m, 2 H, **NCH$_2$CH$_2$C*H*$_2$**) ppm.

ESI-MS m/z: 483.3 $[M]^+$.

B_{12}-linker: β-(3-aminopropyl)cobalamin



$C_{65}H_{97}CoN_{14}O_{14}PCl$
M = 1423.93

β-(3-aminopropyl)cobalamin (**B_{12}-linker**) was prepared according to the literature.[94] Hydroxocobalamin (0.20 g, 0.14 mmol) was dissolved in 10 mL water and subsequently 3-aminopropylchloride HCl (37 mg, 0.29 mmol) and NH_4Cl (77 mg, 1.44 mmol) were added. The solution was degassed with $N_{2(g)}$ for 20 min, followed by addition of zinc dust (94 mg, 1.44 mmol). The reaction mixture was stirred for 5 h under nitrogen atmosphere and filtrated (Whatman filter paper). The filtrate was transferred to a Waters C-18 Sep-Pak cartridge (10 g; activated with 60 mL MeOH and 100 mL H_2O, successively). After washing with 100 mL H_2O, the product **B_{12}-linker** was eluted with 20 mL MeOH. The solvent was removed under reduced pressure and dried by lyophilization. The crude was dissolved in 2 mL MeOH and added to 100 mL rapidly stirring 1:1 DCM/Et_2O. The crystalline product was filtrated, washed with 3x5 mL of each DCM, acetone, and Et_2O and dried *in vacuo* (164 mg, yield: 82 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 7.15 (s, 1 H, **B12-7N**), 7.05 (s, 1 H, **B12-2N**), 6.35 (s, 1 H, **B12-4N**), 6.19 (d, *J* = 3.1 Hz, 1 H, **B12-1R**), 6.05 (s, 1 H, **B12-10**), 4.68 (m, 1 H, **B12-3R**), 4.35 (m, 1 H, **B12-176**), 4.28 (d, *J* = 9.1 Hz, 1 H, **B12-3**), 4.19 -4.12 (m, 1 H, **B12-2R**, 1 H, **B12-4R**, 1 H, **B12-19**), 3.90, 3.87 (dd, *J* = 12.5 Hz, 3.0 Hz, 1 H, **B12-5R**), 3.76, 3.73 (dd, *J* = 12.4 Hz, 4.2 Hz, 1 H, **B12-5R**), 3.60, 3.54 (m, 1 H, **B12-8**, 1 H, **B12-175**), 3.27 (m, 1 H, **B12-13**), 2.99 (m, 1 H, **B12-175**), 2.78 (m, 1 H, **B12-18**), 2.65, 2.61 (m, 2 H, **B12-181**, 2 H, **B12-L3**), 2.55, 2.53 (s, 3 H, **B12-151**, s, 3 H, **B12-51**), 2.59 - 2.30 (m, 2 H, **B12-32**, 2 H, **B12-132**, 1 H, **B12-171**,1 H, **B12-172**, 1 H, **B12-21**, 1 H, **B12-71**), 2.26 (m, 1 H, **B12-21**, 3 H, **B12-10N**, 3 H, **B12-11N**), 2.12 (m, 2 H, **B12-131**, 2 H, **B12-31**, 1 H, **B12-171**, 1 H, **B12-71**, 1 H, **B12-81**), 1.80 (m, 3 H, **B12-7A**, 1 H, **B12-172**, 1 H, **B12-82**), 1.47 (s, 3 H, **B12-12A**), 1.37 (m, 3 H, **B12-2A**, 3 H, **B12-17B**), 1.31 (m, 1 H, **B12-82**), 1.22 (d, *J* = 6.4 Hz, 3 H, **B12-177**), 1.15 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.69 (m, 1 H, **B12-L2**), 0.48 (s, 3 H, **B12-1A**), 0.43 (m, 1 H, **B12-L1**), -0.06 (m, 1 H, **B12-L1**) ppm.

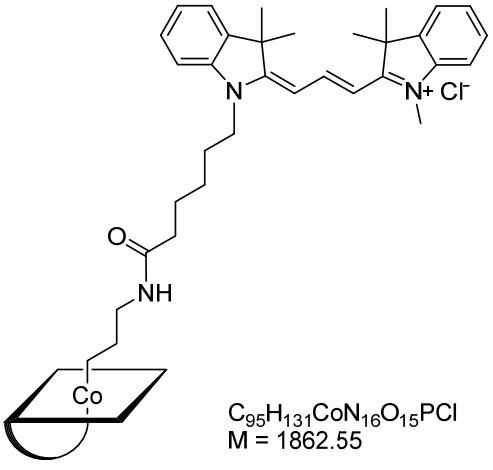
$^{13}C\{^{1}H\}$-NMR (125 MHz, MeOD-d_4, 300 K): δ = 177.93, 177.66, 177.22 (**B12-133**, **B12-33**, **B12-83**), 177.22 (**B12-4**), 177.12 (**B12-16**), 176.24, 175.84 (**B12-22**, **B12-182**), 175.11 (**B12-11**), 174.93, 174.85 (**B12-173**, **B12-72**), 172.49 (**B12-9**), 165.29 (**B12-14**), 164.97 (**B12-6**), 143.30 (**B12-2N**), 139.93 (**B12-9N**), 134.15 (**B12-6N**), 132.21, 132.10 (**B12-5N**, **B12-8N**), 119.68 (**B12-4N**), 111.60 (**B12-7N**), 107.26, 105.05 (**B12-5**, **B12-15**), 95.15 (**B12-10**), 87.62 (**B12-1R**), 86.82 (**B12-1**), 83.29, 83.23 (**B12-4R**), 75.59, 75.55, 75.50 (**B12-19**, **B12-3R**), 73.57, 73.52 (**B12-176**), 70.97 (**B12-2R**), 62.76 (**B12-5R**), 59.32 (**B12-17**), 57.23 (**B12-3**), 56.68 (**B12-8**), 54.32 (**B12-13**), 50.96 (**B12-7**), 48.07 (**B12-12**), 47.55 (**B12-2**), 46.36 (**B12-175**), 45.18 (**B12-71**), 44.04 (**B12-21**), 40.56 (**B12-18**), 39.53 (**B12-L3**), 36.65, 35.76 (**B12-32**, **B12-132**), 33.46, 33.40 (**B12-172**, **B12-82**), 32.68, 32.62, 32.49 (**B12-12B**, **B12-181**, **B12-L1**, **B12-171**), 29.31 (**B12-131**), 27.69, 27.23 (**B12-81, B12-31**), 21.87 (**B12-1A**), 21.66 (**B12-12A**), 20.77, 20.33 (**B12-10N**, **B12-11N**), 20.55 (**B12-7A**), 20.03, 19.99 (**B12-177**), 18.23, 17.85 (**B12-2A**, **B12-17B**), 16.71, 16.57 (**B12-151**, **B12-51**) ppm.

ESI-MS m/z: 1387.7 $[M]^+$, 694.4 $[M+H]^{2+}$.

HPLC: t_R = 7.0 min, 87 % purity.

UV/Vis (H_2O): λ = 266, 281, 349, 375, 526 nm.

B_{12}-Cy3: [Cy3]-β-(3-amidopropyl)cobalamin



$C_{95}H_{131}CoN_{16}O_{15}PCl$
M = 1862.55

Variant A: *In situ* preparation of Cy3 *N*-hydroxysuccinimidyl ester was done by addition of **Cy3** (25 mg, 0.05 mmol) and *N,N'*-disuccinimidyl carbonate (21 mg, 0.08 mmol) to a solution of DIPEA (20 µL) in dry DMF (1.5 mL) under nitrogen atmosphere. The reaction mixture was stirred at 60 °C for 2 h. **B_{12}-linker** (56 mg, 0.04 mmol) and *N,N*-diisopropylethylamine (10 µL) were added and the mixture was stirred for 12 h at room temperature. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Cy3** was dried *in vacuo* (24 mg; yield over two steps: 32 %).

Variant B: Commercially available Cy3 *N*-hydroxysuccinimidyl ester (20 mg, 0.03 mmol) and **B_{12}-linker** (57 mg, 0.04 mmol) were added to DIPEA (15 µL) in DMF (1 mL). The reaction mixture was stirred for 24 h at room temperature under nitrogen atmosphere. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under

reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Cy3** was dried *in vacuo* (30 mg; yield 47 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 8.54 (t, $J_{a,b} = J_{b,c}$ = 13.5 Hz, 1 H, **b**), 7.54 (m, 2 H, **4**, **4'**), 7.45 (m, 2 H, **6**, **6'**), 7.34 (m, 4 H, **7**, **7'**, **5**, **5'**), 7.13 (s, 1 H, **B12-7N**), 7.04 (s, 1 H, **B12-2N**), 6.47, 6.44 (dd, $J_{a,b} = J_{b,c}$ = 13.6 Hz, 2.4 Hz, 2 H, **a**, **c**), 6.35 (s, 1 H, **B12-4N**), 6.17 (d, *J* = 3.1 Hz, 1 H, **B12-1R**), 6.04 (s, 1 H, **B12-10**), 4.66 (m, 1 H, **B12-3R**), 4.35 (m, 1 H, **B12-176**, 1 H, **B12-3**), 4.19 (m, 1 H, **B12-2R**), 4.11 (m, 1 H, **B12-4R**, 1 H, **B12-19**, 2 H, **NC*H*$_2$**), 3.89, 3.87 (dd, *J* = 12.6 Hz, 2.9 Hz, 1 H, **B12-5R**), 3.75, 3.72 (dd, *J* = 12.4 Hz, 4.2 Hz, 1 H, **B12-5R**), 3.69 (s, 3 H, **NC*H*$_3$**), 3.59 (m, 1 H, **B12-8**, 1 H, **B12-175**), 3.22 (m, 1 H, **B12-13**), 2.98 (m, 1 H, **B12-175**), 2.78 (m, 1 H, **B12-18**), 2.67 (m, 2 H, **B12-181**, 2 H, **B12-L3**), 2.54, 2.49 (s, 3 H, **B12-151**, s, 3 H, **B12-51**), 2.54 - 2.45 (m, 2 H, **B12-32**, 2 H, **B12-132**, 1 H, **B12-171**,1 H, **B12-172**, 1 H, **B12-21**, 1 H, **B12-71**), 2.25 (m, 1 H, **B12-21**, 3 H, **B12-10N**, 3 H, **B12-11N**), 2.05 (m, 2 H, **B12-131**, 2 H, **B12-31**, 1 H, **B12-171**, 1 H, **B12-71**, 1 H, **B12-81**, 2 H, **COC*H*$_2$**), 1.82 (m, 3 H, **B12-7A**, 1 H, **B12-172**, 2 H, **NCH$_2$C*H*$_2$**), 1.76 (s, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**), 1.70 (m, 1 H, **B12-82**), 1.56 (m, 2 H, **COCH$_2$C*H*$_2$**), 1.45 (s, 3 H, **B12-12A**), 1.41 (m, 2 H, **NCH$_2$CH$_2$C*H*$_2$**), 1.36 (m, 6 H, **B12-2A**, **B12-17B**, 1 H, **B12-82**), 1.20 (d, *J* = 6.3 Hz, 3 H, **B12-177**), 1.06 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.49 (s, 3 H, **B12-1A**), 0.43 (m, 1 H, **B12-L1**), -0.10 (m, 1 H, **B12-L1**) ppm.

^{13}C{^{1}H}-NMR (125 MHz, MeOD-d_4, 300 K): δ = 177.91, 177.74, 177.19 (**B12-133**, **B12-33**, **B12-83**), 177.10 (**B12-4**,), 176.90 (**B12-16**), 176.66 (**Cy3-2'**), 176.23, 175.92, 176.04, 175.38 (**B12-22**, **B12-182**, **Cy3-2**, **Cy3-*C*O**), 174.94, 174.89 (**B12-173**, **B12-72**), 174.65 (**B12-11**), 172.19 (**B12-9**), 165.24 (**B12-14**), 165.05 (**B12-6**), 152.09 (**Cy3-b**), 144.10, 143.39 (**Cy3-7a**, **7a'**), 143.28 (**B12-2N**), 142.18, 142.09 (**Cy3-3a**, **3a'**), 139.93 (**B12-9N**), 134.04 (**B12-6N**), 132.14 (**B12-5N**, **B12-8N**), 130.04, 130.02 (**Cy3-6**, **6'**), 126.83, 126.74 (**Cy3-5**, **5'**), 123.55, 123.41 (**Cy3-4**, **4'**), 119.76 (**B12-4N**), 112.45, 112.36 (**Cy3-7**, **7'**), 111.52 (**B12-7N**), 106.78, 105.14 (**B12-5**, **B12-15**), 103.88, 103.66 (**Cy3-a**, **c**), 95.37 (**B12-10**), 87.60 (**B12-1R**), 86.80 (**B12-1**), 83.31, 83.25 (**B12-4R**), 75.78 (**B12-19**), 75.61 (**B12-3R**), 73.52, 73.47 (**B12-176**), 70.98 (**B12-2R**), 62.80 (**B12-5R**), 59.32 (**B12-17**), 57.18 (**B12-3**), 56.12 (**B12-8**), 54.40 (**B12-13**), 51.00 (**B12-7**), 50.62, 50.61 (**Cy3-3**, **3'**), 48.11 (**B12-12**), 47.53 (**B12-2**), 46.33 (**B12-175**), 45.07 (**Cy3-NCH$_2$**), 43.86 (**B12-71**), 43.56 (**B12-21**), 40.53 (**B12-18**), 38.85 (**B12-L3**), 36.84 (**Cy3-COCH$_2$**), 36.69, 35.70 (**B12-32**, **B12-132**), 33.71 (**B12-172**), 33.46 (**B12-82**), 32.76, 32.71, 32.65 (**B12-12B**, **B12-181**, **B12-L1**), 32.50 (**B12-171**), 31.86 (**Cy3-NCH$_3$**), 29.42 (**B12-131**), 28.35, 28.34, 28.27, 28.20, 28.16 (**Cy3-3-*C*H$_3$, 3'-*C*H$_3$**, **Cy3-NCH$_2$*C*H$_2$**), 27.74 (**B12-81, B12-31**), 27.47 (**Cy3-NCH$_2$CH$_2$*C*H$_2$**), 26.86 (**Cy3-COCH$_2$*C*H$_2$**), 21.98 (**B12-1A**), 21.30 (**B12-12A**), 20.75, 20.32 (**B12-10N**, **B12-11N**), 20.13 (**B12-7A**), 20.01, 19.97 (**B12-177**), 18.38, 17.68 (**B12-2A**, **B12-17B**), 16.68, 16.62 (**B12-151**, **B12-51**) ppm.

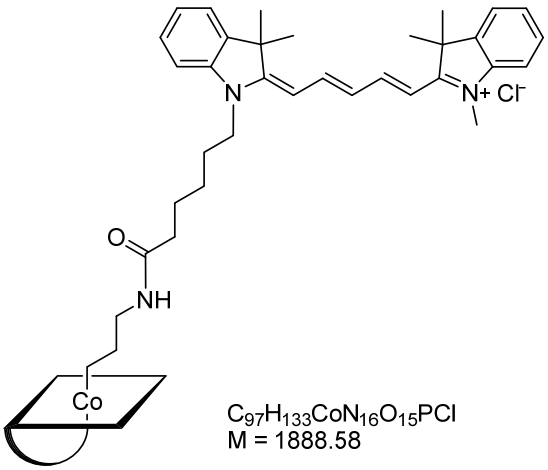
ESI-MS m/z: 1825.9 $[M]^+$, 924.9 $[M+Na]^{2+}$.

HR-ESI-MS m/z: calcd. for $C_{96}H_{135}CoN_{16}O_{16}PNa^{2+}$ ($[M+Na+MeOH]^{2+}$) 940.96159, found 940.94318.

HPLC: t_R = 15.3 min, 98 % purity.

UV/Vis (H_2O): λ = 269, 279, 315, 346, 378, 514, 545 nm.

B_{12}-Cy5: [Cy5]-β-(3-amidopropyl)cobalamin



$C_{97}H_{133}CoN_{16}O_{15}PCl$
M = 1888.58

Variant A: Preparation of Cy5 *N*-hydroxysuccinimidyl ester was done by addition of **Cy5** (72 mg, 0.14 mmol) and *N,N'*-disuccinimidyl carbonate (53 mg, 0.21 mmol) to a mixture of anhydrous pyridine (0.1 mL) in dry DMF (4 mL) under nitrogen atmosphere. The reaction mixture was stirred for 2 h at 60 °C. After the solvent was removed under reduced pressure, the crude product was purified by column chromatography (silica gel, DCM/Hex/MeOH). Cy5 *N*-hydroxysuccinimidyl ester (20 mg, 0.03 mmol) and **B_{12}-linker** (40 mg, 0.03 mmol) were added to a mixture of DIPEA (16 µL), DMF (1 mL) and water (4 mL). The reaction mixture was stirred overnight at room temperature. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Cy5** was dried *in vacuo* (14 mg; yield over two steps: 26 %).

Variant B: Commercially available Cy5 *N*-hydroxysuccinimidyl ester (20 mg, 0.03 mmol) and **B_{12}-linker** (54 mg, 0.04 mmol) were added to DIPEA (15 µl) in DMF (1 mL). The reaction mixture was stirred for 24 h at room temperature under nitrogen atmosphere. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Cy5** was dried *in vacuo* (24 mg; yield 40 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 8.25 (t, $J_{a,b} = J_{b,c} = J_{c,d} = J_{d,e}$ = 13.0 Hz, 2 H, **b**, **d**), 7.50 (m, 2 H, **4**, **4'**), 7.41 (m, 2 H, **6**, **6'**), 7.28 (m, 4 H, **7**, **7'**, **5**, **5'**), 7.13 (s, 1 H, **B12-7N**), 7.04 (s, 1 H, **B12-2N**), 6.62 (t, $J_{b,c} = J_{c,d}$ = 12.4 Hz, 1 H, **c**), 6.34 (s, 1 H, **B12-4N**), 6.29, 6.27 (dd, $J_{a,b} = J_{d,e}$ = 13.7 Hz, 7.7 Hz, 2 H, **a**, **e**), 6.17 (d, J = 2.9 Hz, 1 H, **B12-1R**), 6.04 (s, 1 H, **B12-10**), 4.67 (m, 1 H, **B12-3R**), 4.34 (m, 1 H, **B12-176**), 4.28 (m, 1 H, **B12-3**), 4.19 (m, 1 H, **B12-**

2R), 4.09 (m, 2 H, **B12-19**, 1 H, **B12-4R**, 2 H, **NC*H*$_2$**), 3.87 (m, 1 H, **B12-5R**), 3.74, 3.72 (dd, *J* = 12.6 Hz, 4.0 Hz, 1 H, **B12-5R**), 3.64 (s, 3 H, **NC*H*$_3$**), 3.59 (m, 1 H, **B12-8**, 1 H, **B12-175**), 3.23 (m, 1 H, **B12-13**), 2.96 (m, 1 H, **B12-175**), 2.78 (m, 1 H, **B12-18**), 2.67 (m, 2 H, **B12-L3**), 2.60 (m, 2 H, **B12-181**), 2.55, 2.50 (s, 3 H, **B12-51**, s, 3 H, **B12-151**), 2.56 - 2.27 (m, 1 H, **B12-171**, 1 H, **B12-172**, 2 H, **B12-21**, 1 H, **B12-71**, 2 H, **B12-32**, 2 H, **B12-132**), 2.25 (m, 6 H, **B12-10N**, **B12-11N**), 2.16 - 1.83 (m, 2 H, **B12-31**, 2 H, **B12-131**, 1 H, **B12-81**, 1 H, **B12-171**, 1 H, **B12-71**, 2 H, **COC*H*$_2$**), 1.83 (m, 3 H, **B12-7A**, 1 H, **B12-172**, 1 H, **B12-82**, 2 H, **NCH$_2$C*H*$_2$**), 1.73 (s, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**,), 1.54 (m, 2 H, **COCH$_2$C*H*$_2$**), 1.45 (s, 3 H, **B12-12A**), 1.38 - 1.33 (m, 3 H, **B12-2A**, 2 H, **NCH$_2$CH$_2$C*H*$_2$**, 3 H, **B12-17B**, 1 H, **B12-82**), 1.20 (m, 3 H, **B12-177**), 1.08 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.50 (m, 3 H, **B12-1A**, 1 H, **B12-L2**), 0.39 (m, 1 H, **B12-L1**), -0.13 (m, 1 H, **B12-L1**) ppm.

^{13}C{^{1}H}-NMR (125 MHz, MeOD-d$_4$, 300 K): δ = 177.89, 177.68, 177.17 (**B12-133**, **B12-33**, **B12-83**), 177.08 (**B12-4**), 176.85 (**B12-16**), 176.08 , 175.78 (**B12-22**, **B12-182**), 175.57 (**Cy5-2'**), 175.37 (**Cy5-*C*O**), 174.85, 174.82 (**B12-173**, **B12-72**), 174.73 (**B12-11**), 174.48 (**Cy5-2**), 172.30 (**B12-9**), 165.27, 165.18 (**B12-14**, **B12-6**), 155.66, 155.42 (**Cy5-b, d**), 144.24, 143.61 (**Cy5-7a**, **7a'**), 143.28 (**B12-2N**), 142.62, 142.57 (**Cy5-3a**, **3a'**), 139.93 (**B12-9N**), 134.07 (**B12-6N**), 132.14 (**B12-5N**, **B12-8N**), 129.78 (2 C, **Cy5-6**, **6'**), 126.58 (**Cy5-c**), 126.39, 126.20 (**Cy5-5**, **5'**), 123.46, 123.32 (**Cy5-4**, **4'**), 119.74 (**B12-4N**), 111.95 (2 C, **Cy5-7**, **7'**), 111.55 (**B12-7N**), 106.83 (**B12-5**), 105.10 (**B12-15**), 104.53, 104.14 (**Cy5-a**, **e**), 95.41 (**B12-10**), 87.58 (**B12-1R**), 86.80 (**B12-1**), 83.32, 83.25 (**B12-4R**), 75.62 (**B12-19**, **3R**), 73.54, 73.49 (**B12-176**), 70.99 (**B12-2R**), 62.79 (**B12-5R**), 59.25 (**B12-17**), 57.21 (**B12-3**), 56.14 (**B12-8**), 54.40 (**B12-13**), 51.03 (**B12-7**), 50.58, 50.51 (**Cy5-3**, **3'**), 48.10 (**B12-12**), 47.52 (**B12-2**), 46.30 (**B12-175**), 44.70 (**Cy5-N*C*H$_2$**), 44.04, 43.76 (**B12-71**, **B12-21**), 40.53 (**B12-18**), 38.89 (**B12-L3**), 36.84 (**Cy5-CO*C*H$_2$**), 36.66, 35.70 (**B12-32**, **B12-132**), 33.53, 33.45 (**B12-172**, **B12-82**), 32.76, 32.68 (**B12-12B**, **B12-181**, **B12-L1**), 32.39 (**B12-171**), 31.58 (**Cy5-NCH$_3$**), 29.40 (**B12-131**), 28.20 (**Cy5-NCH$_2$*C*H$_2$**), 27.97, 27.78 (**3-*C*H$_3$, 3'-*C*H$_3$**), 27.73 (**B12-31**, **B12-81**), 27.51 (**Cy5-NCH$_2$CH$_2$*C*H$_2$**), 26.88 (**Cy5-COCH$_2$*C*H$_2$**), 21.97 (**B12-1A**), 21.39 (**B12-12A**), 20.76, 20.31 (**B12-10N**, **B12-11N**), 20.21 (**B12-7A**), 20.02, 19.97 (**B12-177**), 18.19, 17.73 (**B12-2A**, **B12-17B**), 16.68, 16.63 (**B12-151**, **B12-51**) ppm.

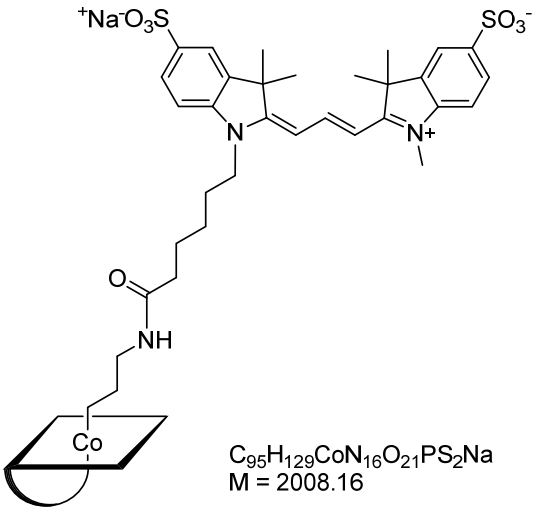
ESI-MS m/z: 1852.9 $[M]^+$, 926.9 $[M+H]^{2+}$.

HR-ESI-MS m/z: calcd. for $C_{98}H_{137}CoN_{16}O_{16}PNa^{2+}$ ($[M+Na+MeOH]^{2+}$) 953.96942, found 953.95183; 942.96065 $[M+H+MeOH]^{2+}$.

HPLC: t_R = 16.5 min, 98 % purity.

UV/Vis (H_2O): λ = 268, 283, 317, 357, 378, 599, 642 nm.

B_{12}-Sulfo-Cy3: [Sulfo-Cy3]-β-(3-amidopropyl)cobalamin



Sulfo-Cy3 *N*-hydroxysuccinimidyl ester (20 mg, 0.03 mmol) and **B_{12}-linker** (45 mg, 0.03 mmol) were added to a solution of DIPEA (15 µl) in DMF (1 mL). The reaction mixture was stirred for 24 h at room temperature under nitrogen atmosphere. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Sulfo-Cy3** was dried *in vacuo* (33 mg; yield: 61 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 8.54 (t, $J_{a,b}$ = $J_{b,c}$ = 13.5 Hz, 1 H, **b**), 7.93 (m, 4 H, **4**, **4'**, **6**, **6'**), 7.42, 7.38 (dd, *J* = 20.8 Hz, 8.3 Hz, 2 H, **7**, **7'**), 7.13 (s, 1 H, **B12-7N**), 7.04 (s, 1 H, **B12-2N**), 6.48, 6.46 (dd, $J_{a,b}$ = $J_{b,c}$ = 13.6 Hz, 1.9 Hz, 2 H, **a**, **c**), 6.35 (s, 1 H, **B12-4N**), 6.17 (d, *J* = 3.0 Hz, 1 H, **B12-1R**), 6.03 (s, 1 H, **B12-10**), 4.67 (m, 1 H, **B12-3R**), 4.35 (m, 1 H, **B12-176**), 4.29 (m, 1 H, **B12-3**), 4.19 - 4.08 (m, 1 H, **B12-2R**, 1 H **B12-19**,1 H, **B12-4R**, m, 2 H, **NC*H*$_2$**), 3.89, 3.87 (dd, *J* = 12.6 Hz, 2.8 Hz, 1 H, **B12-5R**), 3.75, 3.72 (dd, *J* = 12.6 Hz, 4.1 Hz, 1 H, **B12-5R**), 3.70 (s, 3 H, **NC*H*$_3$**), 3.56 (m, 2 H, **B12-8, B12-175**), 3.31 (m, 1 H, **B12-13**), 3.03 (m, 1 H, **B12-175**), 2.78 (m, 1 H, **B12-18**), 2.69 - 2.35 (m, 2 H, **B12-L3**, 2 H, **B12-181**, 1 H, **B12-172**, 1 H, **B12-171**, 1 H, **B12-71**, 2 H, **B12-32**, 2 H, **B12-132**, 2 H, **B12-21**), 2.55, 2.49 (s, 3 H, **B12-51**, s, 3 H, **B12-151**), 2.25 (m, 3 H, **B12-10N**, 3 H, **B12-11N**), 2.16 - 2.01 (m, 2 H, **B12-31**, 2 H, **B12-131**, 1 H, **B12-71**, 1 H, **B12-171**, 1 H, **B12-81**, 2 H, **COC*H*$_2$**), 1.83 (m, 1 H, **B12-172**, 3 H, **B12-7A** 2 H, **NCH$_2$C*H*$_2$**), 1.77 (s, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**, 1 H, **B12-82**), 1.55 (m, 2 H, **COCH$_2$C*H*$_2$**),1.44 (s, 3 H, **B12-12A**), 1.37-1.29 (m, 3 H, **B12-2A**, 2 H, **NCH$_2$CH$_2$C*H*$_2$**, 3 H, **B12-17B**, 1 H, **B12-82**), 1.21 (d, *J* = 6.3 Hz, 3 H, **B12-177**), 1.05 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.53 (m, 1 H, **B12-L2**), 0.48 (s, 3 H, **B12-1A**), 0.29 (m, 1 H, **B12-L1**), -0.20 (m, 1 H, **B12-L1**) ppm.

^{13}C{^{1}H}-NMR (125 MHz, MeOD-d_4, 300 K): δ = 177.86, 177.73, 177.20 (**B12-133**, **B12-33**, **B12-83**), 177.45, 177.11, 176.91, 176.88, 176.10, 175.86, 175.04 (**Cy3-2'**, **B12-4**, **B12-16**, **Cy3-2**, **B12-22**, **B12-182**, **Cy3-CO**), 174.92, 174.86 (**B12-173**, **B12-72**), 174.73 (**B12-11**), 172.30 (**B12-9**), 165.26 (**B12-14**), 165.12 (**B12-6**), 152.69 (**Cy3-b**), 145.35, 144.75 (**Cy3-5**, **5'**), 144.09, 144.12 (**Cy3-7a**, **7a'**), 143.24 (**B12-2N**), 142.13, 142.12 (**Cy3-3a**, **3a'**), 139.94 (**B12-9N**), 134.03 (**B12-6N**), 132.15, 132.12 (**B12-5N**, **B12-8N**), 128.22 (**Cy3-6**, **6'**), 121.46, 121.37 (**Cy3-4**, **4'**), 119.76 (**B12-4N**), 112.33, 112.05 (**Cy3-7**, **7'**), 111.49 (**B12-7N**), 106.90 (**B12-5**), 105.10 (**B12-15**), 104.83, 104.79 (**Cy3-a**, **c**), 95.38 (**B12-10**), 87.56 (**B12-1R**), 86.78

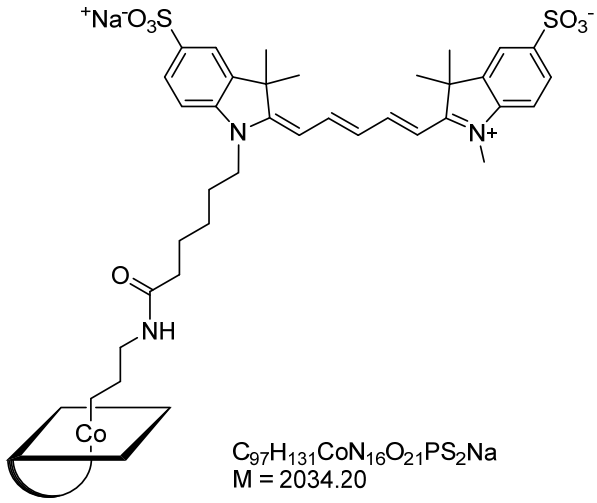
(**B12-1**), 83.25, 83.19 (**B12-4R**), 75.51 (**B12-3R, B12-19**), 73.54 (**B12-176**), 71.01 (**B12-2R**), 62.72 (**B12-5R**), 59.24 (**B12-17**), 57.14 (**B12-3**), 56.17 (**B12-8**), 54.45 (**B12-13**), 51.01 (**B12-7**), 50.76, 50.68 (**Cy3-3**, **3'**), 48.04 (**B12-12**), 47.52 (**B12-2**), 46.25 (**B12-175**), 45.20 (**Cy3-NCH$_2$**), 43.95, 43.75 (**B12-71**, **B12-21**), 40.43 (**B12-18**), 38.86 (**B12-L3**), 36.64, 36.53, 35.72 (**Cy3-COCH$_2$**, **B12-32, B12-132**), 33.57, 33.52 (**B12-172**, **B12-82**), 32.77, 32.57, 32.43 (**B12-12B, B12-181**, **B12-L1**, **B12-171**), 32.07 (**Cy3-NCH$_3$**), 29.34 (**B12-131**), 28.27, 28.27, 28.25, 28.14, 28.08 (**Cy3-3-CH$_3$, 3'-CH$_3$, Cy3-NCH$_2$CH$_2$**), 27.77, 27.70, 27.03 (**Cy3-NCH$_2$CH$_2$CH$_2$**, **B12-81**, **B12-31**), 26.32 (**Cy3-COH$_2$CH$_2$**), 21.96 (**B12-1A**), 21.40 (**B12-12A**), 20.75, 20.32 (**B12-10N**, **B12-11N**), 20.19 (**B12-7A**), 20.01, 19.97 (**B12-177**), 18.28, 17.65 (**B12-2A**, **B12-17B**), 16.74, 16.61 (**B12-151**, **B12-51**) ppm.

HR-ESI-MS m/z: calcd. for $C_{95}H_{128}CoN_{16}O_{21}PS^{2-}$ ([M-H]$^{2-}$) 991.89978, found 991.89869.

HPLC: t_R = 10.6 min, 92 % purity.

UV/Vis (H_2O): λ = 268, 281, 291, 317, 343, 376, 518, 551 nm.

B_{12}-Sulfo-Cy5: [Sulfo-Cy5]-β-(3-amidopropyl)cobalamin



$C_{97}H_{131}CoN_{16}O_{21}PS_2Na$
M = 2034.20

Sulfo-Cy5 *N*-hydroxysuccinimidyl ester (20 mg, 0.03 mmol) and **B_{12}-linker** (34 mg, 0.03 mmol) were added to a solution of DIPEA (15 µL) in DMF (1 mL). The reaction mixture was stirred for 8 h at room temperature under nitrogen atmosphere. Completeness of the reaction was indicated by analytical HPLC analysis. The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-Sulfo-Cy5** was dried *in vacuo* (36 mg; yield: 71 %).

^{1}H-NMR (500 MHz, MeOD-d$_4$, 300 K): δ = 8.31 (m, $J_{a,b} = J_{b,c} = J_{c,d} = J_{d,e}$ = 13.0 Hz, 2 H, **b**, **d**), 7.87 (m, 4 H, **4**, **4'**, **6**, **6'**), 7.36 - 7.30 (m, 2 H, **7**, **7'**), 7.13 (s, 1 H, **B12-7N**), 7.05 (s, 1 H, **B12-2N**), 6.66 (t, $J_{b,c} = J_{c,d}$ = 12.3 Hz, 1 H, **c**), 6.36 (s, 1 H, **B12-4N**), 6.33, 6.30 (dd, $J_{a,b} = J_{d,e}$ = 13.6 Hz, 5.9 Hz, 2 H, **a**, **e**), 6.17 (d, J = 3.1 Hz, 1 H, **B12-1R**), 6.04 (s, 1 H, **B12-10**), 4.68 (m, 1 H, **B12-3R**), 4.35 (m, 1 H, **B12-176**), 4.28 (m, 1 H, **B12-3**), 4.19 (m, 1 H, **B12-2R**), 4.11 (m, 1 H, **B12-19**, 1 H, **B12-4R**, 2 H, **NC*H$_2$***), 3.89, 3.87 (dd, J = 12.7 Hz, 2.9 Hz, 1 H, **B12-5R**), 3.74, 3.72 (dd, J = 12.6 Hz, 4.1 Hz, 1 H, **B12-5R**), 3.65 (s, 3 H, **NC*H$_3$***), 3.59 (m, 1 H, **B12-8**, 1 H, **B12-175**), 3.31 (m, 1 H, **B12-13**), 3.03 (m, 1 H, **B12-175**), 2.79 (m 1 H, **B12-18**), 2.67 (m, 2 H, **B12-L3**), 2.60 (m, 2 H, **B12-181**), 2.55, 2.50 (m, 3 H, **B12-51**, 3 H, **B12-151**), 2.59 -

2.36 (m, 1 H, **B12-172**, 1 H, **B12-171**, 1 H, **B12-71**, 2 H, **B12-32**, 2 H, **B12-132**, 2 H, **B12-21**), 2.25 (m, 3 H, **B12-10N**, 3 H, **B12-11N**), 2.14 - 1.89 (m, 2 H, **B12-31**, 2 H, **B12-131**, 1 H, **B12-71**, 1 H, **B12-171**, 1 H, **B12-81**, 2 H, **COC*H*$_2$**), 1.83 (m, 3 H, **B12-7A**, 1 H, **B12-172**, 2 H, **NCH$_2$C*H*$_2$**), 1.75, 1.73 (m, 12 H, **3-C*H*$_3$, 3'-C*H*$_3$**, 1 H, **B12-82**), 1.53 (m, 2 H, **COCH$_2$C*H*$_2$**), 1.45 (s, 3 H, **B12-12A**), 1.38, 1.34 (m, 3 H, **B12-2A**, 2 H, **NCH$_2$CH$_2$C*H*$_2$**, 3 H, **B12-17B**, 1 H, **B12-82**), 1.21 (d, *J* = 6.3 Hz, 3 H, **B12-177**), 1.06 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.51 (m, 1 H, **B12-L2**), 0.49 (s, 3 H, **B12-1A**), 0.31 (m, 1 H, **B12-L1**), -0.17 (m, 1 H, **B12-L1**) ppm.

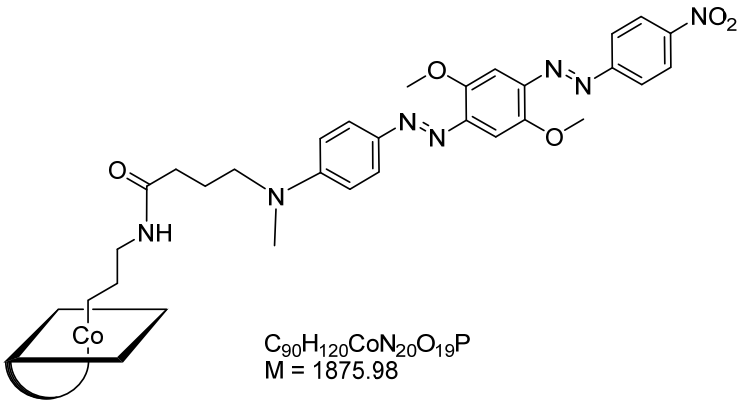
^{13}C{^{1}H}-NMR (125 MHz, MeOD-d$_4$, 300 K): δ = 177.84, 177.72, 177.18 (**B12-133**, **B12-33**, **B12-83**), 177.05 (**B12-4**), 176.85 (**B12-16**), 176.16 (**Cy5-2'**), 176.10, 175.87 (**B12-22**, **B12-182**), 175.15, 175.11 (**Cy5-2**, **Cy5-*C*O**), 174.92, 174.86 (**B12-173**, **B13-72**), 174.70 (**B12-11**), 172.25 (**B12-9**), 165.23, 165.09 (**B12-14**, **B12-6**), 156.34, 156.09 (**Cy5-b, d**), 145.55, 145.05, 143.47, 143.32 (**Cy5-5**, **5'**, **Cy5-7a**, **7a'**), 143.24 (**B12-2N**), 142.51, 142.49 (**Cy5-3a, 3a'**), 139.94 (**B12-9N**), 134.00 (**B12-6N**), 132.12, 132.10 (**B12-5N**, **B12-8N**), 128.00, 127.68 (2 C, **Cy5-6**, **6'**, **Cy5-c**), 121.28, 121.22 (**Cy5-4**, **4'**), 119.74 (**B12-4N**), 111.73 (**B12-7N**), 111.53, 111,48 (2 C, **Cy5-7**, **7'**), 106.85 (**B12-5**), 105.41, 105.18 (**Cy5-a**, **e**), 105.11 (**B12-15**), 95.36 (**B12-10**), 87.55 (**B12-1R**), 86.79 (**B12-1**), 83.23, 83.16 (**B12-4R**), 75.58, 75.53 (**B12-19**, **3R**), 73.51, 73.47 (**B12-176**), 71.00 (**B12-2R**), 62.68 (**B12-5R**), 59.25 (**B12-17**), 57.21 (**B12-3**), 56.16 (**B12-8**), 54.46 (**B12-13**), 50.99 (**B12-7**), 50.56, 50.51 (**Cy5-3**, **3'**), 49.83 (**B12-12**), 48.04 (**B12-2**), 46.28 (**B12-175**), 44.81 (**Cy5-N*C*H$_2$**), 43.92, 43.62 (**B12-71**, **B12-21**), 40.45 (**B12-18**), 38.82 (**B12-L3**), 36.61, 36.58 (**B12-32**, **B12-132**), 35.70 (**Cy5-CO*C*H$_2$**), 33.61, 33.51 (**B12-172**, **B12-82**), 32.66, 32.58 (**B12-12B**, **B12-181**, **B12-L1**), 32.41 (**B12-171**), 31.78 (**Cy5-N*C*H$_3$**), 29.32 (**B12-131**), 28.01 (**Cy5-NCH$_2$*C*H$_2$**), 27.88, 27.87, 27.69, 27.68 (**3-CH$_3$**, **3'-CH$_3$**, **B12-31**, **B12-81**), 27.14 (**Cy5-NCH$_2$CH$_2$*C*H$_2$**), 26.43 (**Cy5-COCH$_2$*C*H$_2$**), 21.95 (**B12-1A**), 21.37 (**B12-12A**), 20.74, 20.30 (**B12-10N**, **B12-11N**), 20.17 (**B12-7A**), 19.99, 19.95 (**B12-177**), 18.30, 17.63 (**B12-2A**, **B12-17B**), 16.71, 16.60 (**B12-151**, **B12-51**) ppm.

HR-ESI-MS m/z: calcd. for $C_{97}H_{130}CoN_{16}O_{21}PS^{2-}$ ([M-H]$^{2-}$) 1004.90760, found 1004.90741.

HPLC: t_R = 11.5 min, 96 % purity.

UV/Vis (H_2O): λ = 267, 281, 319, 332, 360, 604, 648 nm.

B_{12}-BHQ-2: [BHQ-2]-β-(3-amidopropyl)cobalamin



$C_{90}H_{120}CoN_{20}O_{19}P$
M = 1875.98

A solution of BHQ-2 *N*-hydroxysuccinimidyl ester (10 mg, 0.02 mmol) in DMSO (0.4 mL) was added to a solution of **B_{12}-linker** (75 mg, 0.04 mmol) in 0.5 mL carbonate-bicarbonate buffer at pH 8.3 (0.1 M). The reaction mixture was stirred overnight at room temperature. Completeness of the reaction was indicated by analytical HPLC analysis (method 1b). The solvent was removed under reduced pressure. The crude reaction mixture was purified by preparative HPLC and the product **B_{12}-BHQ-2** was dried *in vacuo* (15 mg; yield: 48 %).

^{1}H-NMR (500 MHz, MeOD-d_4, 300 K): δ = 8.40 (d, *J* = 9.1 Hz, 2 H, **BHQ-Q15**), 8.08 (d, *J* = 9.1 Hz, 2 H, **BHQ-Q14**), 7.89 (d, *J* = 9.2 Hz, 2 H, **BHQ-Q7**), 7.53, 7.47 (s, 1 H, **BHQ-Q10**, s, 1 H, **BHQ-Q10'**), 7.12 (s, 1 H, **B12-7N**), 7.03 (s, 1 H, **B12-2N**), 6.84 (d, *J* = 9.3 Hz, 2 H, **BHQ-Q6**), 6.34 (s, 1 H, **B12-4N**), 6.15 (d, *J* = 3.0 Hz, 1 H, **B12-1R**), 5.99 (s, 1 H, **B12-10**), 4.67 (m, 1 H, **B12-3R**), 4.33 (m, 1 H, **B12-176**), 4.25 (m, 1 H, **B12-3**), 4.18 (m, 1 H, **B12-2R**), 4.09 (m, 1 H, **B12-19**, 1 H, **B12-4R**), 4.06 (s, 3 H, **BHQ-OMe**), 4.01 (s, 3 H, **BHQ-OMe**), 3.87, 3.85 (dd, *J* = 12.6 Hz, 2.9 Hz, 1 H, **B12-5R**), 3.73, 3.71 (dd, *J* = 12.6 Hz, 4.2 Hz, 1 H, **B12-5R**), 3.57 (m, 2 H, **B12-8**, **B12-175**), 3.45 (m, 2 H, **BHQ-Q4**), 3.16 (m, 1 H, **B12-13**), 3.09 (s, 3 H, **BHQ-NMe**), 2.98 (m, 1 H, **B12-175**), 2.77 (m, 1 H, **B12-18**), 2.71 (m, 2 H, **B12-L3**), 2.58 (m, 2 H, **B12-181**), 2.56 - 2.36 (m, 1 H, **B12-172**, 2 H, **B12-32**, 2 H, **B12-132,** 1 H, **B12-171**, 1 H, **B12-71**, 2 H, **B12-21**), 2.53, 2.47 (s, 3 H, **B12-51**, s, 3 H, **B12-151**), 2.25, 2.24 (s, 3 H, **B12-10N**, 3 H, **B12-11N**), 2.09 (m, 2 H, **B12-131**, 1 H, **B12-31**, 1 H, **B12-171**, 1 H, **B12-71**, 2 H, **BHQ-Q2**), 1.93 (m, 1 H, **B12-81**), 1.83 (m, 2 H, **BHQ-Q3**, 3 H, **B12-7A**, 1 H, **B12-31**, 1 H, **B12-172**), 1.70 (m, 1 H, **B12-82**), 1.41 (s, 3 H, **B12-12A**), 1.38 (s, 3 H, **B12-2A**), 1.32 (m, 3 H, **B12-17B**, 1 H, **B12-82**, 1 H, **B12-L2**), 1.19 (d, *J* = 6.2 Hz, 3 H, **B12-177**), 1.06 (m, 3 H, **B12-12B**, 1 H, **B12-81**), 0.49 (m, 1 H, **B12-L2**, s, 3 H, **B12-1A**), 0.42 (m, 1 H, **B12-L1**), -0.12 (m, 1 H, **B12-L1**) ppm.

^{13}C{^{1}H}-NMR (125 MHz, MeOD-d_4, 300 K): δ = 177.84, 177.70, 177.10, 177.05, 176.82, 176.09, 175.77 (**B12-133**, **B12-33**, **B12-83**, **B12-22**, **B12-182**, **B12-4**, **B12-16**), 174.91, 174.87, 174.82, 174.74 (**B12-173**, **B12-172**, **B12-11**, **BHQ-Q1**), 172.27 (**B12-9**), 165.19, 165.18 (**B12-14**, **B12-6**), 157.65 (**BHQ-Q13**), 154.95 (**BHQ-Q11/11'**), 153.74 (**BHQ-Q5**), 152.21 (**BHQ-Q11/11'**), 149.95 (**BHQ-Q16**), 147.85 (**BHQ-Q12/9**),145.53 (**BHQ-Q8**), 143.38 (**BHQ-Q12/9**), 143.24 (**B12-2N**), 139.92 (**B12-9N**), 134.07 (**B12-6N**), 132.16, 132.12 (**B12-**

5N, **B12-8N**), 127.29 (**BHQ-Q7**), 125.83 (**BHQ-Q15**), 124.69 (**BHQ-Q14**), 119.74 (**B12-4N**), 112.73 (**BHQ-Q6**), 111.52 (**B12-7N**), 106.91, 105.10 (**B12-5**, **B12-15**), 102.09, 101.17 (**BHQ-Q10'**, **BHQ-Q10**), 95.41 (**B12-10**), 87.57 (**B12-1R**), 86.80 (**B12-1**), 83.26, 83.21 (**B12-4R**), 75.61 (**B12-19**), 75.57, 75.54 (**B12-3R**), 73.51, 73.47 (**B12-176**), 71.01 (**B12-2R**), 62.72 (**B12-5R**), 59.23 (**B12-17**), 57.27, 57.22, 57.17 (**B12-3**, **BHQ-OMe**), 56.31 (**B12-8**), 54.48 (**B12-13**), 52.66 (**BHQ-Q4**), 51.11 (**B12-7**), 48.09 (**B12-12**), 47.52 (**B12-2**), 46.29 (**B12-175**), 44.15 (**B12-71**), 43.76 (**B12-21**), 40.53 (**B12-18**), 39.02 (**BHQ-NMe**), 38.97 (**B12-L3**), 36.62, 35.77 (**B12-32**, **B12-132**), 34.11 (**BHQ-Q2**), 33.53 (**B12-172**), 33.41 (**B12-82**), 32.76, 32.72, 32.67 (**B12-12B**, **B12-181**, **B12-L1**), 32.39 (**B12-171**), 29.37 (**B12-131**), 27.71, 27.65 (**B12-81**, **B12-31**), 25.19 (**B12-L2**), 24.50 (**BHQ-Q3**), 21.97 (**B12-1A**), 21.30 (**B12-12A**), 20.75, 20.32 (**B12-10N**, **B12-11N**), 20.21 (**B12-7A**), 20.01, 19.97 (**B12-177**), 18.19 (**B12-17B**), 17.72 (**B12-2A**), 16.65, 16.60 (**B12-151**, **B12-51**) ppm.

ESI-MS m/z: 1875.8 $[M+H]^+$, 1898.7 $[M+Na]^+$.

HPLC: t_R = 18.0 min, 99 % purity.

UV/Vis (H_2O): λ = 266, 280, 321, 351, 559 nm.

5.3.3 Fluorescence Quantum Yields and Steady-State Anisotropy

The procedure for the determination of the fluorescence quantum yields and static fluorescence anisotropy of the free dyes and the dyes attached to B_{12} follows the established techniques. All spectra were recorded at 22 °C. Cresyl violet perchlorate (Φ = 0.54) and rhodamine 6G (Φ = 0.94) in ethanol were used as references.[17] All samples were dissolved in *PBS buffer*. Stock solutions were diluted to different concentrations chosen such that the absorption at the excitation wavelength (500 nm Cy3, 580 nm Cy5) did not exceed 0.1. Absorbance and fluorescence spectra were recorded. The quantum yield was calculated according to the following equation:

$$\Phi_s = \Phi_r \frac{I_s}{I_r}\frac{OD_r}{OD_s}\frac{\eta_s^2}{\eta_r^2} \quad (4)$$

s and r refer to the sample and reference, respectively. Φ is the fluorescence quantum yield, I is the integrated fluorescence intensity, OD is the optical density at the excitation wavelength and η is the refractive index of the solvent.

For static fluorescence anisotropy measurements an automated polarizer was used to record fluorescence intensities of polarized excitation and emission in vertical (v, 0°) and horizontal (h, 90°) positions.

Stability measurements were done exciting a solution of B_{12}-Cy3 in *PBS buffer* at 532 nm for 10 min. The absorbance and emission was measured before and after irradiation.

5.3.4 Preparation of RNA (*btuB* Riboswitch of *E.coli*)

Plasmid Isolation

The *btuB* riboswitch sequences of *Escherichia coli* pPC4 (270 nts, Appendix) designed by Choudhary *et al.*,[176] and pSG1 (202 nts, Appendix) designed by Gallo *et al.*, [87] are encoded on plasmids derived from the pRZ-vector.[180] Glycerol stocks with cells (*E. coli*, DH5α) carrying the plasmid (pPC4 or pSG1) were used to inoculate an LB culture containing the appropriate amount of ampicillin (100 mg/mL) and incubated overnight at 37 °C, 210 rpm. Purification was done using a PureYield Plasmid Maxiprep System (Promega).

Transcription of pPC4 and pSG1

In order to obtain a linear DNA template all plasmids were digested with *EcoR*1 at the restriction site located downstream of the riboswitch region. After purification the plasmids were transcribed by *in vitro* transcription using self-made T7 RNA polymerase according to the literature.[87,181] More specifically, 15 µg/mL *EcoR*1 digested plasmid, 20 mM $MgCl_2$ (pPC4) or 30 mM $MgCl_2$ (pSG1), respectively, 5 mM of each NTP (ATP, CTP, GTP, and UTP), and 20 µL/mL T7 RNA polymerase were mixed in 1x *Transcription Buffer*. The reaction was allowed to proceed for 5-6 h at 37 °C and shaking (300 rpm). For pSG1 the $MgCl_2$ concentration was increased to 40 mM and a thermal cycling method was applied (2 min 90 °C, 5 min 65 °C, 10 min 37 °C; repeated three times) to cleave the hammerhead ribozyme at the 5'- and the hepatitis delta virus ribozyme at the 3'- flanking regions of the riboswitch sequence. Transcription products were purified via ethanol precipitation, denaturing PAGE (10 % w/v, 7 M urea), and eluted by electroelution. After ethanol precipitation the RNA was dissolved in water and concentrated by Vivaspin centrifugation. The RNA concentration was determined spectroscopically from the absorbance at 260 nm using the extinction coefficients of the RNA nucleobases yielding 0.6 nmol RNA/mL transcription.

5.3.5 Fluorescence Native Gels

Fluorescence native gel electrophoresis was performed with the *btuB* riboswitch RNA (pPC4) and **DNA2-Cy5** (5'-Sulfo-Cy5-ACGTAGTCCGAAATATAT-3') or **PNA1S-Cy5** (Lys-tataaagcctgatg-spacer-Sulfo-Cy5, synthesis described in Chapter 5.4.1). Samples containing pPC4 (3 µM) and **DNA1-Cy5** (3 µM) or **PNA1S-Cy5** (3 µM), were denatured in 50 mM Tris-HCl (pH 7.5), 100 mM KCl for 60 s at 90 °C. $MgCl_2$ (20 mM) was added and the RNA was

allowed to prefold for 15 min at 37 °C (smFRET conditions). An equal amount of glycerol was added and the samples were loaded onto a 6 % (w/v) gel containing 3 mM MgOAc, 66 mM HEPES, and 34 mM Tris-HCl at pH 7.4. Electrophoresis was performed at 4 °C using a running buffer of the same conditions as described. Fluorescence bands of Cy5 in the gels were detected on a Typhoon scanner.

5.3.6 In-line Probing Experiments

Prior to in-line probing experiments pPC1 RNA (Appendix) which carries a 5'-triphosphate was dephosphorylated. The RNA pPC1 or pSG1, respectively, was 5'-end labeled with [^{32}P]-γ-ATP for a sensitive detection. In-line probing titration experiments were carried out using the [^{32}P]-5'-labeled RNA in the final concentration noted below for each experiment. The RNA was incubated in the dark for 40 h at 25°C in 50 mM Tris-HCl (pH 8.3), 100 mM KCl, and 20 mM $MgCl_2$ in presence or absence of different cobalamin derivatives. The cleaved bands were separated by 10 % (w/v, 7 M urea) denaturing PAGE, visualized by phosphor imaging, and analyzed with ImageQuantTL. For every experiment ladders and controls were used.[49,87]

5'-Labeling of RNA

The RNA pPC1 (480 pmol) was dephosphorylated with 0.16 U/µL Thermosensitive Alkaline Phosphatase (TSAP) for 30 min at 37 °C, and purified by chloroform-phenol extraction.[176] pPC1 or pSG1 (240 pmol), respectively, in ME buffer were 5'-labled with [^{32}P]-γ-ATP (15 µCi/µL) by addition of T4 polynucleotide kinase (PNK, 1 U/µL) in PNK buffer. The reaction was allowed to proceed for 30 min at 37 °C and shaking (300 rpm). The RNA was purified via denaturing PAGE (8 % w/v, 7 M urea), visualized by phosphor imaging, eluted with *Elution buffer*, and concentrated by ethanol precipitation. After dissolving the RNA in water the concentration was determined by scintillation counting.

Ladders and Controls

The alkaline hydrolysis cleavage ladder (OH^-) was generated by incubation of the 5'-labeled RNA (same concentration as used in corresponding experiment) in 50 mM Na_2CO_3 and 1 mM EDTA at pH 9.0 for 5 min at 90 °C.

The T1 cleavage ladder (T1) which cleaves the RNA at each guanine residue was generated by incubation of the 5'-labeled RNA (same concentration as used in corresponding experiment) and ribonuclease T1 (100 mU) in 25 mM sodium citrate at pH 5.0 for 5 min at 55 °C.

The uncleaved 5'-labeled RNA (same concentration as used in corresponding experiment) is used as unreacted control (c). A negative control (–) was generated by incubation without cobalamin and a positive control (+) by addition of 10 µM AdoCbl to the in-line probing experiment under the above described conditions.

Titration Experiments

Titration experiments were performed using pPC1 RNA at a final concentration of 38 nM of which 5 nM were ^{32}P-labeled under the described conditions (*vide supra*) and addition of:

B_{12}-Cy3: 0.3 µM, 1.2 µM, 9.2 µM, 27.7 µM, 55.3 µM, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 0.7 mM, 1.0 mM, 1.5 mM.

B_{12}-Sulfo-Cy3: 0.3 µM, 1.2 µM, 9.4 µM, 28.1 µM, 56.3 µM, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 0.7 mM, 1.0 mM, 1.5 mM.

B_{12}-Cy5: 0.3 µM, 2.0 µM, 7.9 µM, 15.9 µM, 31.8 µM, 63.5 µM, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 0.7 mM, 0.9 mM.

B_{12}-Sulfo-Cy5: 0.3 µM, 1.9 µM, 7.5 µM, 15.0 µM, 30.0 µM, 60.0 µM, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 0.7 mM, 0.9 mM.

B_{12}-linker: 0.3 µM, 3.4 µM, 13.6 µM, 27.2 µM, 54.3 µM, 0.1 mM, 0.2 mM, 0.3 mM, 0.5 mM, 0.7 mM, 1.0 mM, 1.5 mM.

Titration experiments were performed with pSG1 RNA at a final concentration of 38 nM of which 3.5 nM were ^{32}P-labeled under the described conditions (*vide supra*) and addition of:

B_{12}-BHQ-2: 1.2 µM, 2.3 µM, 4.7 µM, 9.4 µM, 18.8 µM, 37.5 µM, 75.0 µM, 0.2 mM, 0.3 mM, 0.6 mM, 1.2 mM, 1.8 mM.

Evaluation of the Titration Experiments

A background correction for intensity changes due to differing loading volumes was done by applying a correction factor (R). R is calculated per lane by the intensity of a distinct constant region relative to the most intense lane.
Intensities of each band were quantified, divided by R and plotted versus the concentration of each B_{12} derivative.

The affinity constant K_a was determined for each characteristic cleavage site by fitting the mean of three independent titration experiments to a 1:1 binding isotherm:

$$I_{rel} = \delta F + (\delta B - \delta F) \cdot \frac{\left([RNA] + [B_{12}] + \frac{1}{K_a}\right) - \sqrt{\left([RNA] + [B_{12}] + \frac{1}{K_a}\right)^2 - 4 \cdot [RNA] \cdot [B_{12}]}}{2 \cdot [RNA]} \quad (5)$$

with δF and δB being intensity of free (F) and bound (B) RNA, and K_a the affinity constant.[97] The final dissociation constant K_D for each derivative was calculated as weighted mean with standard deviation from the single cleavage sites that could be fitted, derived from at least three independent experiments. Cleavage sites 1, 2b, 3, 4 and 8 were used in **B_{12}-Cy3** experiments, cleavage sites 1 and 8 in **B_{12}-Cy5** experiments, and cleavage sites 1, 2, 2b, 3, 4 and 8 in **B_{12}-BHQ-2** experiments. For **B_{12}-Sulfo-Cy3**, **B_{12}-Sulfo-Cy5** and **B_{12}-linker** no K_D values could be calculated.

Control Experiments

Overview gels were performed with pPC1 RNA at a final concentration of 38 nM of which 5 nM were ^{32}P-labeled under the described conditions (*vide supra*) and addition of:

B_{12}-Cy3, **B_{12}-Sulfo-Cy3**: 0.1, 0.5, 1.5 mM; **B_{12}-Cy5**, **B_{12}-Sulfo-Cy5**: 0.1, 0.5, 0.8 mM.

Control experiments were performed with pSG1 RNA at a final concentration of 38 nM of which 3.5 nM were ^{32}P-labeled under the described conditions (*vide supra*) and addition of:

Cy3, **Cy5:** 0.3 µM, 0.5 mM, 1.0 mM; **Sulfo-Cy3**, **Sulfo-Cy5**: 0.3 µM, 0.7 mM, 1.5 mM.

5.4 Methods Chapter 3 (Labeling of RNA)

The following chapter is based on the manuscript accepted for publication:

"Internal Labeling Strategy of Large RNAs with Minimal Perturbation Using Fluorescent PNA"

by Anita G. Schmitz, Susann Zelger-Paulus, Gilles Gasser, and Roland K. O. Sigel in **ChemBioChem**, 2015.[169]

5.4.1 PNA Synthesis

General Procedure for Solid Phase PNA Synthesis (SPPS)

The procedure for SPPS was performed according to Anstaett *et al.*[170] More specifically, the reaction was carried out manually in 5 mL polypropylene one-way syringes as reaction vessels, which were equipped with a frit at the bottom. They were filled with 100 mg of polystyrene resin beads TentaGel® S RAM-Lys(Boc)Fmoc (0.23 mmol/g). The resin was swollen in dimethylformamide (DMF) for 1 h before use. All reactions were performed on a mechanical shaker with 600 rpm, soaking approximately 2.5–3.5 mL of freshly prepared solutions into the syringe. Fluorenylmethoxycarbonyl (Fmoc)/benzhydryloxycarbonyl (Bhoc) protected PNA monomers (5 equiv.) were pre-activated in Eppendorf tubes before each coupling step for 5 min with *O*-(7-azobenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU) (4.5 equiv.) in DMF, adding *N,N*-diisopropylethylamine (DIPEA) and 2,6-lutidine (10 equiv. each) under sonication. For each coupling step the resin beads were treated with the activated acid (1.5 h) and subsequently washed with DMF and dichloromethane (DCM). The coupling step was monitored with the Kaiser test.[158,159] Substances other than the usual PNA monomers were coupled in the same way, unless noted otherwise. Double Fmoc deprotection was performed with piperidine in DMF (2:8, v/v; 2 min, 10 min). The resin beads were then washed successively with DMF, dichloromethane (DCM) and DMF. The whole procedure (deprotection, coupling, monitoring) was repeated for every monomer until the sequence was completed. The resin was then shrunk with methanol (30 min) and dried under vacuum. Final cleavage of the PNA oligomer from the resin and the deprotection of all Bhoc and *tert*-butyloxycarbonyl (Boc) side chain protecting groups were simultaneously performed in trifluoroacetic acid (TFA)/triisopropylsilane (TIS)/H_2O (38:2:1, v/v/v; 3x 2 h). Following the removal of TFA under reduced pressure, the crude product was purified by preparative HPLC (method 4). MALDI-MS spectra, and LC-MS (method 3) were recorded and the purities were determined from the HPLC UV traces at 260 nm.

HPLC method 3 (analytical): column Agilent Zorbax SB300-C18; flow-rate 0.5 mL/min; solvent A: 0.1 % formic acid in water, solvent B: ACN; gradient: 0-3 min 5 % B, 3-17 min 5-100 % B, 17-20 min 100 % B, 20-21 min 100-5 % B, 21-25 min 5 % B.

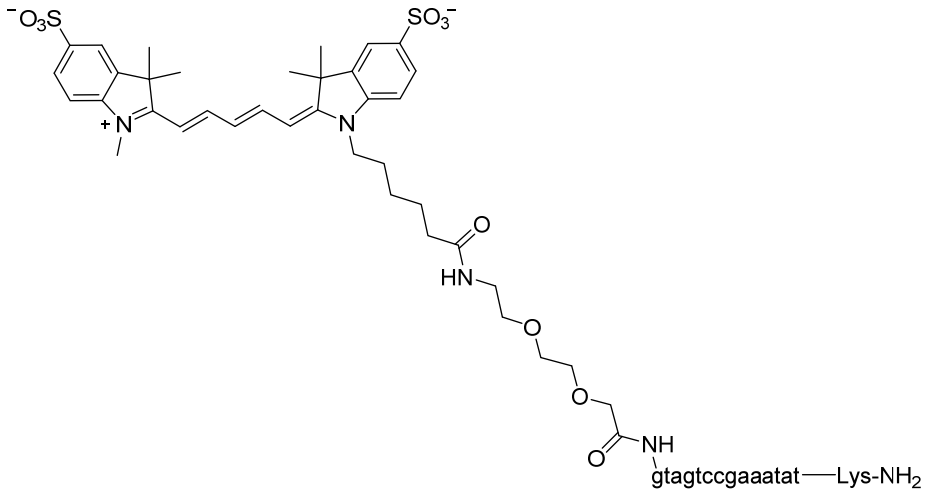
HPLC method 4 (preparative): column Agilent Zorbax SB300-C18; flow-rate 21 mL/min; solvent A: 0.1 % TFA in water, solvent B: ACN; gradient: 0-21 min 5-45 % B, 21-25 min 45-100 % B, 25-30 min 100 % B, 30-32 min 100-0 % B.

In the following PNA sequences are denoted *N*→*C* terminus in small letters.

Precursor PNA1S: Fmoc-spacer-gtagtccgaaatat-Lys-resin

SPPS was performed according to the general procedure (*vide supra*), until the last coupling of the PNA sequence was completed. After deprotection a Fmoc-AEEA-OH spacer (5 equiv.) was coupled to the sequence.

PNA1S-Cy5: Sulfo-Cy5-spacer-gtagtccgaaatat-Lys-NH_2



For the subsequent coupling of the dye Sulfo-Cy5 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA1S** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{196}H_{249}N_{88}O_{52}S_2^+$ ($[M+2H]^+$) 4730.9, found 4730.9.

ESI-MS m/z: 1184.0 (6, $[M+5H]^{4+}$), 947.7 (71, $[M+6H]^{5+}$), 789.9 (100, $[M+7H]^{6+}$), 677.2 (27, $[M+8H]^{7+}$), 592.6 (2, $[M+9H]^{8+}$).

HPLC: t_R = 8.8 min, 95 % purity.

PNA1S-Cy3: Sulfo-Cy3-spacer-gtagtccgaaatat-Lys-NH_2

For the subsequent coupling of the dye Sulfo-Cy3 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA1S** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

In an alternative approach the resin with the PNA sequence was saturated with Kiton Red S (14 equiv.) in DMF/DIPEA 30:1 for 30 min prior to the addition of Sulfo-Cy3.[171]

MALDI-MS: calcd. for $C_{194}H_{247}N_{88}O_{52}S_2^+$ ($[M+2H]^+$) 4704.9, found 4704.7.

ESI-MS m/z: 942.0 (3, $[M+6H]^{5+}$), 785.3 (100, $[M+7H]^{6+}$), 673.3 (36, $[M+8H]^{7+}$), 587.9 (4, $[M+9H]^{8+}$).

HPLC (method 3, flow-rate 0.3 mL/min): t_R = 10.8 min, 95 % purity.

Precursor PNA2: Fmoc-gtagtccgaa-Lys-resin

SPPS was performed according to the general procedure (*vide supra*) until the last coupling of the PNA sequence was completed.

PNA2-Cy5: Sulfo-Cy5-gtagtccgaa-Lys-NH_2

For the subsequent coupling of the dye Sulfo-Cy5 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA2** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{146}H_{184}N_{65}O_{37}S_2^+$ ($[M+2H]^+$) 3503.4, found 3503.4.

ESI-MS m/z: 1169.3 (9, $[M+4H]^{3+}$), 877.2 (53, $[M+5H]^{4+}$), 702.0 (100, $[M+6H]^{5+}$), 585.1 (32, $[M+7H]^{6+}$).

HPLC: t_R = 8.7 min, 95 % purity.

PNA2S-Cy5: Sulfo-Cy5-spacer-gtagtccgaa-Lys-NH_2

Fmoc-AEEA-OH spacer (5 equiv.) was coupled to ½ of precursor **PNA2** according to the general SPPS procedure. For the subsequent coupling of the dye Sulfo-Cy5 NHS (1.5 equiv.) in DMF/DIPEA 30:1 was added and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{152}H_{195}N_{66}O_{40}S_2^+$ ([M+2H]$^+$) 3648.5, found 3648.5.

ESI-MS m/z: 1217.3 (2, [M+4H]$^{3+}$), 913.5 (35, [M+5H]$^{4+}$), 731.1 (100, [M+6H]$^{5+}$), 609.4 (44, [M+7H]$^{6+}$).

HPLC: t_R = 8.8 min, 95 % purity.

Precursor PNA3: Fmoc-acgtagtccg-Lys-resin

SPPS was performed according to the general procedure (*vide supra*) until the last coupling of the PNA sequence was completed.

PNA3-Cy5: Sulfo-Cy5-acgtagtccg-Lys-NH_2

For the subsequent coupling of the dye Sulfo-Cy5 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA3** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{145}H_{184}N_{63}O_{38}S_2^+$ ([M+2H]$^+$) 3479.4, found 3479.3.

ESI-MS m/z: 1161.2 (6, [M+4H]$^{3+}$), 871.2 (47, [M+5H]$^{4+}$), 697.2 (100, [M+6H]$^{5+}$), 581.1 (25, [M+7H]$^{6+}$).

HPLC: t_R = 8.6 min, 100 % purity.

PNA3S-Cy5: Sulfo-Cy5-spacer-acgtagtccg-Lys-NH_2

Fmoc-AEEA-OH spacer (5 equiv.) was coupled to ½ of precursor **PNA3** according to the general SPPS procedure. For the subsequent coupling of the dye Sulfo-Cy5 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{151}H_{195}N_{64}O_{41}S_2^+$ ([M+2H]$^+$) 3622.5, found 3624.4.

ESI-MS m/z: 1209.4 (5, [M+4H]$^{3+}$), 907.4 (44, [M+5H]$^{4+}$), 726.2 (100, [M+6H]$^{5+}$), 605.3 (44, [M+7H]$^{6+}$), 519.0 (4, [M+8H]$^{7+}$).

HPLC: t_R = 8.8 min, 97 % purity.

Precursor PNA4: Fmoc-accaagagcg-Lys-resin

SPPS was performed according to the general procedure (*vide supra*) until the last coupling of the PNA sequence was completed.

PNA4-Cy3: Sulfo-Cy3-accaagagcg-Lys-NH_2

For the subsequent coupling of the dye Sulfo-Cy3 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA4** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{143}H_{180}N_{69}O_{34}S_2^+$ ($[M+2H]^+$) 3471.4, found 3471.3.

ESI-MS m/z: 1158.5 (8, $[M+4H]^{3+}$), 869.1 (38, $[M+5H]^{4+}$), 695.6 (100, $[M+6H]^{5+}$), 579.8 (71, $[M+7H]^{6+}$), 497.2 (5, $[M+8H]^{7+}$).

HPLC: t_R = 8.2 min, 95 % purity.

5.4.2 Preparation of RNA (Group II Intron Ribozyme)

Polymerase Chain Reaction PCR for D135-L1sh4, and D135-L14sh[182]

The truncated version D135-L14 (638 nts) of the group II intron *Sc*.ai5γ from *Saccharomyces cerevisiae* is encoded on the plasmid named pT7D135-L14, designed by Steiner *et al.*[42] D135-L14 was modified by the deletion of 9 bp (5'-AAUUAAUAA-3') in loop L1 and 9 bp (5'-AAUAUAUUU-3') in loop L4 using site-directed mutagenesis PCR giving the plasmids pSPA7a and pSPA7b.

More specifically, the PCR reactions were carried out using 0.04 U/µL Phusion High-Fidelity DNA polymerase (*Phu*). For the deletion in loop L1 300 ng of the initial plasmid pT7D135-L14, 0.25 µM each forward primer P1 and reverse primer P2 (Table 8), and 200 µM of each dNTP were mixed with *Phu* polymerase in 1x *Phu* buffer. The method described in Table 9 was applied. The methylated DNA was digested with *Dpn*I (0.2 U/µL) for 1 h at 37 °C. Afterwards, the plasmid pSPA7a was transformed into chemically competent cells (*Escherichia coli* XL1-Blue) via heat shock and further isolated by using the PureYield Plasmid Miniprep System (Promega) following the manufactures instructions.

Table 8: Primer sequences used for PCR.

Name	Sequence
P1	5'-GTCTGATATTCAATTATTATCCCACGCTCTTGGTAGGGATAAAATGG-3'
P2	5'-CCATTTTATCCCTACCAAGAGCGTGGGATAATAATTGAATATCAGAC-3'
P3	5'-GCTAACTTATACTATAGGTGATACCCACGGACTACGTAGGGTATTGCG-3'
P4	5'-CGCAATACCCTACGTAGTCCGTGGGTATCACCTATAGTATAAGTTAGC-3'

Table 9: PCR program.

Reaction	time	temperature	
Initial denaturation	5 min	95 °C	
Denaturation	45 s	95 °C	20 cycles
Primer annealing	45 s	55 °C	
Elongation	3 min	72 °C	
Final elongation	10 min	72 °C	

The deletion in loop L4 was performed using pSPA7a (300 ng), 0.25 µM each forward primer P3 and reverse primer P4 (Table 8), 200 µM of each dNTP, *Phu* polymerase (0.04 U/µl) and 1x *Phu* buffer. The method described in Table 9 was applied. Digestion of the reaction mixture was done with *Dpn*I (0.2 U/µL) for 1 h at 37 °C. The plasmid was inserted into chemically competent cells (*E. coli* XL1-Blue) via heat shock transformation and further isolated by using the PureYield Plasmid Miniprep System (Promega) following the manufactures instructions.

Plasmid Isolation

Cells carrying the plasmid (pT7D135, pT7D135-L14, pSPA7a, or pSPA7b) were used to inoculate an LB culture containing the appropriate amount of ampicillin (100 mg/mL) and incubated overnight at 37 °C, 210 rpm. Purification was done using a PureYield Plasmid Maxiprep System (Promega).

Transcription of D135, D135-L14, D135-L1shL4, and D135-L14sh

In order to obtain a linear DNA template all plasmids were digested with *Hind*III (10 U/µL) at the restriction site located downstream of the intron region. The purified plasmids were transcribed by *in vitro* transcription using self-made T7 RNA polymerase according to the literature.[42,181] More specifically, 7.5 µg/mL *Hind*III digested plasmid, 25 mM $MgCl_2$, 5 mM of each NTP (ATP, CTP, GTP, and UTP), and 20 µL/mL T7 RNA polymerase were mixed in 1x *Transcription Buffer*. The reaction was allowed to proceed for 5-6 h at 37 °C and shaking

(300 rpm). Transcription products were purified via ethanol precipitation, denaturing PAGE (5 % w/v, 7 M urea), and eluted by electroelution. The RNA was concentrated by ethanol precipitation and dissolved in water. The RNA concentration was determined spectroscopically from the absorbance at 260 nm using the extinction coefficients of the single RNA nucleobases yielding 0.75 nmol RNA/mL transcription.

5.4.3 UV Melting Experiments

UV melting experiments of RNA-PNA and RNA-DNA hybrids were undertaken in order to obtain the respective melting temperature T_m. All complementary RNA strands were extended by three uracil residues on the 3'- and 5'-end (Table 10) to avoid terminal group effects. The PNA concentration was determined in *PBS buffer* (0.1 M phosphate buffer, 0.1 M NaCl, pH 7.4) at 260 nm and 70 °C using the extinction coefficients of the PNA nucleobases ($\varepsilon_{PNA,A}$ = 13700 $M^{-1}cm^{-1}$, $\varepsilon_{PNA,G}$ = 11700 $M^{-1}cm^{-1}$, $\varepsilon_{PNA,C}$ = 6600 $M^{-1}cm^{-1}$, $\varepsilon_{PNA,T}$ = 8600 $M^{-1}cm^{-1}$).[152] Solutions of **PNA1S-Cy3**, **PNA1S-Cy5**, **PNA2-Cy5**, **PNA2S-Cy5**, **PNA3-Cy5**, **PNA3S-Cy5**, **DNA1-Cy3** or **DNA2-Cy5** together with the complementary **RNA1**–**RNA5** were used at a concentration of 2 µM in degassed *Reaction Buffer*. Prior to the experiment the duplexes were denatured at 90 °C for 1 min and then cooled down to room temperature. The thermal program was applied by heating to 90 °C, cooling to 10 °C, both at 0.5 °C/min with a starting and returning temperature of 90 °C. The procedure was repeated three times.[172] The absorption was recorded at 260 nm, and the data was analyzed by the derivative method.

Table 10: RNA sequences used for UV melting experiments with PNA-dye or DNA-dye derivatives.

Name	Sequence	Loop	PNA/DNA
RNA1	5'-UUUAUUAAUAACGCUCUUGGUUUU-3'	L1	**DNA1-Cy3**
RNA2	5'-UUUCGCUCUUGGUUUU-3'	L1	**PNA4-Cy3**
RNA3	5'-UUUAUAUAUUUCGGACUACGUUUU-3'	L4	**DNA2-Cy5**
RNA4	5'-UUUAUAUUUCGGACUACUUU-3'	L4	**PNA1S-Cy3, PNA1S-Cy5, PNA2-Cy5, PNA2S-Cy5**
RNA5	5'-UUUCGGACUACGUUUU-3'	L4	**PNA3-Cy5, PNA3S-Cy5**

5.4.4 Fluorescence Native Gels

Fluorescence native gel electrophoresis was performed according to the literature.[42] More specifically, samples containing the ribozyme RNA D135-L14 (2 µM), **DNA1-Cy3** (2 or 4 µM, as indicated), and **DNA2-Cy5** (2 or 4 µM, as indicated) were denatured in *Reaction Buffer* for 45 s at 90 °C. $MgCl_2$ (100 mM) was added and the RNA was allowed to prefold for 20 min at

42 °C (smFRET conditions). For experiments with PNA-dye oligonucleotides the procedure was modified. Samples containing the ribozyme D135-L14, D135-L14sh or D135-L1shL4 (2 µM) were denatured in *Reaction Buffer* for 45 s at 90 °C, followed by addition of $MgCl_2$ (100 mM) and prefolding for 10 min at 42 °C. **PNA3-Cy5** (2 µM), and **PNA4-Cy3** (2 µM) were added and the samples were incubated for 10 min at 42 °C. An equal amount of glycerol was added and the samples were loaded onto a 5 % (w/v) gel containing 3 mM MgOAc, 66 mM HEPES, and 34 mM Tris-HCl at pH 7.4. Electrophoresis was performed at 4 °C using a running buffer of the same conditions as described. Fluorescence bands of Sulfo-Cy3 and Sulfo-Cy5 in the gels were detected on a Typhoon scanner. Additionally, RNA bands were stained with ethidium bromide (EtBr) and detected by UV excitation at 260 nm.

5.4.5 Fluorescence Quantum Yields and Steady-State Anisotropy

The procedure for the determination of the fluorescence quantum yields and static fluorescence anisotropy of the free dyes and the dyes attached to the DNA or PNA oligonucleotides follows the established techniques (see Chapter 5.3.3).[17] All spectra were recorded at 22 °C. Cresyl violet perchlorate (Φ = 0.54) and Rhodamine 6G (Φ = 0.94) in ethanol were used as references.[17] All samples were dissolved in water or *Reaction Buffer* with 100 mM $MgCl_2$. Stock solutions were diluted to different concentrations chosen such that the absorption at the excitation wavelength did not exceed 0.1. Absorbance and fluorescence spectra were recorded.

5.4.6 Single-Turnover Cleavage Assays

Single-turnover cleavage assays were performed according to the literature.[42] More specifically, the ribozyme substrate called 17/7 (5'-CGUGGUGGGACAUUUUC▼GAGCGGU-3') contains the last 17 nts of the 5'-exon sequence (including the intron binding sites IBS1 and IBS2) and the first 7 nts of the intron including the 5'-splice site. 17/7 was 5'-labeled with [^{32}P]-γ-ATP and T4 polynucleotide kinase using standard procedures. DNA-dye-oligonucleotides were used in 6 fold excess over RNAs (D135, D135-L14, D135L1shL4, and D135-L14sh). PNA-dye oligonucleotides (**PNA3-Cy5** and **PNA4-Cy3**) were used in a 1:1 ratio of RNA:PNA. **DNA1-Cy3**, **DNA2-Cy5** and DNA-biotin (1.2 µM) were heat-annealed to the ribozyme (200 nM) at 90 °C for 45 s in *Reaction Buffer* followed by $MgCl_2$ addition (100 mM) and incubation at 42 °C for 20 min. In experiments with PNA, DNA-biotin (1.2 µM) was heat-annealed to the ribozyme (200 nM) at 90 °C for 45 s in *Reaction Buffer* followed by $MgCl_2$ addition (100 mM) and incubation at 42 °C for 10 min. Afterwards **PNA4-Cy3** and **PNA3-Cy5** (200 nM) were added and the sample was incubated at 42 °C for 10 min. ^{32}P-labeled 17/7 (2 nM) was separately incubated in *Reaction Buffer* and $MgCl_2$ (100 mM) and then added to biotinylated and Cy3/Cy5-labeled ribozyme in a 1:1 ratio to start the cleavage

reaction. Aliquots (1 µL) were removed from the reaction mixture at specific time points. The cleaved bands were separated on 18 % (w/v, 7 M urea) denaturing polyacrylamide gels, visualized by phosphor imaging and analyzed with ImageQuantTL. Reaction rate constants were determined by quantification of the gels and fitting to a single exponential expression:

$$frac\ (product) = (1 - A_1) - A_2 \times e^{-kt} \quad (6)$$

with A_1 being the fraction of uncleaved substrate, A_2 the fraction of product formed, and *k* the first order rate constant in min^{-1}. Values of k_{obs} are calculated as arithmetic mean with standard deviation from at least three independent experiments.

5.4.7 Single-Molecule FRET Experiments

Single-Molecule FRET measurements were conducted at room temperature according to the procedure by Steiner *et al.*[42] More specifically, samples containing the ribozyme construct D135-L14 or D135-L1shL4 (0.5 µM), and DNA-biotin (2.5 µM) for surface immobilization were denatured in *Reaction Buffer* containing Trolox ((±)-6-Hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid) for 45 s at 90 °C. $MgCl_2$ (100 mM) was added and samples were allowed to prefold for 10 min at 42 °C. PNA-Cy3 (**PNA4-Cy3**) and PNA-Cy5 (**PNA1S-Cy5**, **PNA2-Cy5**, **PNA2S-Cy5**, **PNA3-Cy5** or **PNA3S-Cy5**) each 0.5 µM were added and the samples were incubated for another 10 min at 42 °C. For control experiments 2.5 µM of each DNA-Cy3 (**DNA1-Cy3**) and DNA-Cy5 (**DNA2-Cy5**) were added to D135-L14 (0.5 µM), and DNA-biotin (2.5 µM) in *Reaction Buffer* containing Trolox, followed by incubation for 45 s at 90 °C. $MgCl_2$ (100 mM) was added and samples were allowed to prefold for 15 min at 42 °C. Each sample was diluted to 15 pM and bound to a BSA-biotin-streptavidin-coated quartz slide surface. Excess oligomers were removed from the slide by a washing step with *Reaction Buffer*. The system was flushed with an oxygen-scavenging system containing 10 % (w/v) glucose, 50 µg/mL glucose oxidase, and 10 µg/mL catalse to reduce photobleaching. All experiments were carried out in home-built flow chambers. The donor (I_D) and acceptor (I_A) fluorescence signals of optically resolved single molecules were recorded. Alternating laser excitation (ALEX) of the green and the red laser was used to check for a constant acceptor signal to clearly distinguish between photobleaching events and actual conformational changes.[183] For every experiment at least 100 dynamic molecules were analyzed. In the data analysis, filter, background (20 darkest pixels), and bleed-through (7 %) corrections were applied. A smoothing function was not used for the evaluated time trajectories. Apparent E_{FRET} values were accumulated in histograms and the distributions were analyzed with multiple Gaussian fits combined with bootstrapping in the software BOBA FRET[175] to reveal reoccurring mean FRET values. Dwell-time analysis was done using the

method vbFRET based on Hidden Markov modeling (HMM) implemented in the software MASH.[179]

5.5 Methods Chapter 4 (Collaborations)

5.5.1 Imaging of an RNA-Splicing Process Using Fluorescent PNA Labels

Synthesis of precursor PNA-Ex1: Fmoc-gtagtaagtc-Lys-resin

SPPS was performed according to the general procedure (Chapter 5.4.1) until the last coupling of the PNA sequence was completed.

Synthesis of PNA-Ex1-Cy3: Sulfo-Cy3-gtagtaagtc-Lys-NH_2

For the subsequent coupling of the dye Sulfo-Cy3-NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of deprotected precursor **PNA-Ex1** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{145}H_{185}N_{64}O_{38}S_2^+$ ($[M+2H]^+$) 3492.4, found 3492.4.

ESI-MS m/z: 1164.9 (8, $[M+4H]^{3+}$), 874.1 (100, $[M+5H]^{4+}$), 699.5 (5, $[M+6H]^{5+}$).

HPLC: t_R = 9.0 min, 100 % purity.

Synthesis of precursor PNA-Ex2: Fmoc-taatacatag-Lys^{Mtt}-resin

To obtain the precursor **PNA-Ex2** 100 mg TentaGel® S RAM (0.24 mmol/g) were used. In a first step 5 equiv. of *N*-α-Fmoc-*N*-ε-4-methyltrityl-L-lysine (Fmoc-Lys(Mtt)-OH)[177] were coupled to the resin followed by coupling of the monomers according to the general procedure (SPPS, Chapter 5.4.1) until the last coupling of the PNA sequence was completed.

Synthesis of PNA-Ex2-Cy5: NH_2-taatacatag-Lys(Sulfo-Cy5)-NH_2

Mtt deprotection of the precursor **PNA-Ex2** was performed in TFA/TIS/DCM (1:5:94, v/v/v) for 2 min and three times repeated including a washing step with DCM. After washing with DMF Sulfo-Cy3 NHS ester (1.5 equiv.) in DMF/DIPEA 30:1 was added to ½ of Mtt and Fmoc deprotected precursor **PNA-Ex2** and shaken for 5 h. After washing with DMF and DCM the coupling step was repeated. The cleavage and purification were performed as described in the general procedure.

MALDI-MS: calcd. for $C_{145}H_{183}N_{64}O_{36}S_2^+$ ($[M+2H]^+$) 3486.4, found 3486.5.

ESI-MS m/z: 1163.5 (5, $[M+4H]^{3+}$), 872.9 (45, $[M+5H]^{4+}$), 698.5 (100, $[M+6H]^{5+}$), 582.3 (31, $[M+7H]^{6+}$).

HPLC: t_R = 8.8 min, 95 % purity.

5.5.2 RNA Interaction with Rhenium Compounds

The interactions of a series of rhenium compounds, which can be activated by light, were tested on the D135-L14 RNA construct. D135-L14 (1 µM) was incubated with increasing concentrations of each rhenium compound **Re-PLPG-NLS**, **Re-PLPG**, or **Re-NH_2** in buffer (50 mM Tris-HCl, 18 mM NaCl, pH 7.2) for 30 min at room temperature and irradiated at 350 nm for 5 min. Control experiments were carried out in the dark. An equal volume of *Denaturing Loading Buffer* was added to each samples and denaturing polyacrylamide gel electrophoresis was performed (5 % w/v, 7 M urea) for 3 h at room temperature in TBE 1x buffer. RNA bands were detected by UV shadowing. Additionally RNA bands were stained with ethidium bromide and detected by UV excitation at 260 nm.[178]

6 Summary

Ribonucleic acid (RNA) is a multifunctional biomolecule, which is essential for living organism as it is involved in countless fundamental processes (Chapter 1).

The focus of this work is on the investigation of fluorescence labeling strategies to elucidate complex folding processes of RNAs, in particular the *btuB* riboswitch of *Escherichia coli*. This RNA possesses a gene regulating function based on a structural rearrangement upon specific binding of its metabolite coenzyme B_{12} (adenosylcobalamin). The folding mechanism of this particular RNA and metabolite binding are not fully clarified yet and their elucidation could give additional details about the functionality of riboswitches and related systems. Many applications could arise from this knowledge, *e.g.* usage as highly selective antibacterial drug targets, or intracellular sensors. Förster Resonance Energy Transfer (FRET) is the technique applied in the herein presented studies to elucidate the RNA folding and metabolite binding. FRET is a distance dependent energy transfer from an excited donor fluorophore to a corresponding acceptor dye. With this technique distance changes between the fluorophores can be monitored by detecting changes of the fluorescence intensities upon energy transfer. Thus, conformational and kinetic information about labeled molecules in motion can be obtained. On the single-molecule level (sm) this method reveals especially detailed information about folding pathways that are hidden in ensemble measurements due to averaging of individual folding events.

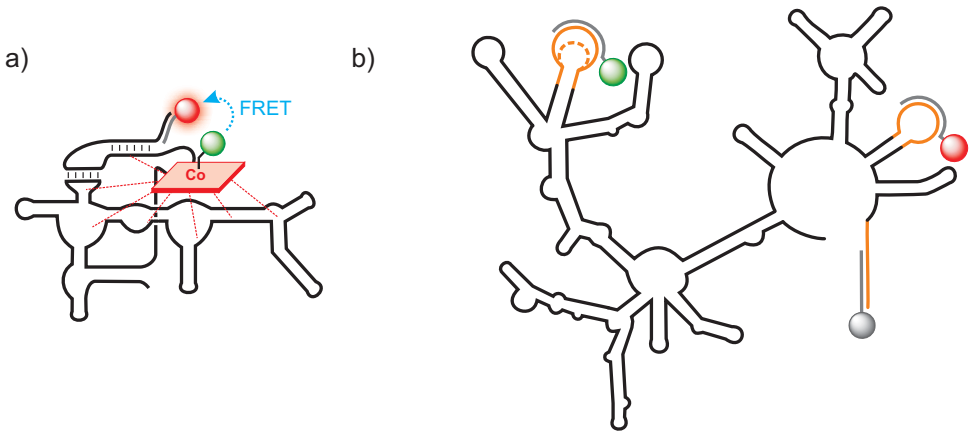


Figure 76: New labeling strategies to investigate the folding of large RNAs via FRET. a) Labeling of the *btuB* riboswitch and its metabolite B_{12} with dyes. b) Labeling of a group II intron ribozyme by hybridization of fluorescent PNA or DNA derivatives.

For fluorescence detection suitable labels need to be present at the molecules of interest. Chapter 2 focuses on the dye labeling of the metabolite to study metabolite binding to the RNA and metabolite induced folding of the RNA (Figure 76 a). In the case of the *btuB* riboswitch the metabolite, cobalamin (B_{12}), is fairly large. B_{12} itself is not fluorescent and can

therefore not be visualized by fluorescence techniques. However, its structure gives the opportunity to fluorescently label the metabolite by direct attachment of dyes. For FRET a suitable dye pair is needed with overlapping emission spectrum of the donor fluorophore and absorption spectrum of the acceptor. The fluorescent dyes of choice are Cy3 (donor) and Cy5 (acceptor), a common FRET pair, which are well established for nucleic acid investigations. Alternatively Cy5 can be replaced by the non-fluorescent 'black hole quencher' BHQ-2. The use of fluorescence quenchers minimizes possible background problems due to fluorescence of the labeled metabolite in high concentrations. The donor dye can be located at the RNA and the acceptor dye at the B_{12}, or vice versa. The axial position of the cobalamin was chosen for the attachment of the dye since this position of the cobalamin has been found to influence only the binding affinity of the B_{12} to the riboswitch but not to alternate the folding behavior of the RNA. Modifications at the corrin ring on the other hand would lead to a changed folding of the RNA. Thus, a linker group was introduced to the cobalt center at the β-side of the B_{12} to which the dyes were coupled via NHS ester chemistry.

An important prerequisite using this methodology is that the modified B_{12} derivatives do not change the riboswitch folding drastically. Therefore, in addition to the synthesis and characterization of novel dye-labeled B_{12} derivatives, their interactions with the *btuB* riboswitch were investigated with in-line probing experiments in comparison to the natural ligand adenosylcobalamin. B_{12} derivatives labeled with sulfonated, and thus negatively charged, cyanine dyes were found to be no suitable probes for FRET experiments with the *btuB* riboswitch. **B_{12}-Sulfo-Cy3** is changing the folding behavior drastically, and **B_{12}-Sulfo-Cy5** does not bind at all to the riboswitch. These results underline the specificity of the riboswitch towards different B_{12} derivatives. As expected, the binding affinities of **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-BHQ-2** are reduced compared to adenosylcobalamin, and are in the range of vitamin B_{12} (~ 200 µM). Furthermore, it was observed that the attachment of the bulky dyes (Cy3, Cy5 or BHQ-2) to the B_{12} is slightly affecting the RNA conformational switch. However, the changes in the riboswitch folding are fairly small. **B_{12}-Cy3**, **B_{12}-Cy5** and **B_{12}-BHQ-2** are, therefore, promising candidates for FRET experiments.

Fluorescence measurements showed that for choosing the measuring conditions in FRET experiments potential photolysis of the Co-C bond has to be considered as it can lead to a loss of the fluorophore. Furthermore, it was observed that B_{12} can quench the fluorescence of the attached dye by itself up to a certain degree.

Nevertheless, this new strategy of fluorescently labeled metabolites gives the opportunity to get insights into the binding mechanism to RNA, which are otherwise inaccessible. In addition, the investigations of the modified B_{12} derivatives with the *btuB* riboswitch revealed, that negative charges at the B_{12} are affecting the folding of the riboswitch.

Chapter 3 deals with the fluorescence labeling of the RNA (Figure 76 b). Site-specific internal labeling methods for large RNAs exist, but they suffer from insufficient labeling yields and an extensive synthesis. Commonly used indirect methods on the other hand require fairly large modifications of the RNA for a hybridization of DNA-fluorophore conjugates, which can lead to misfolding of the RNA. Based on this known scheme, we improved the labeling strategy by using Peptide Nucleic Acid (PNA) instead of DNA. PNA is an uncharged non-natural nucleic acid analog, which is known to bind stronger to RNA than DNA does. Thus, modifications of the RNA can be minimized to bind the label.

New PNA derivatives carrying sulfonated Cy3 or Cy5 FRET dyes, linked directly or via an additional spacer, were synthesized by solid phase synthesis and then characterized. The dye-conjugates were tested on a well-studied RNA model system, which allows a direct comparison of the previously used labeling scheme to the improved one. To this end, we chose a truncated version of the mitochondrial group II intron *Sc.*ai5γ derived from *Saccharomyces cerevisiae* (D135-L14, Figure 76 b). It is a catalytic active large RNA molecule (> 600 nts), which has already been successfully characterized by smFRET. Therefore potential interference between the PNA-labels and the RNA folding would easily be detected.

The fluorescence properties of the samples are among the effects that have to be considered for an optimal fluorescence labeling. An important aspect is their dependency on environmental influences. Fluorescence quantum yields and steady-state anisotropy measurements demonstrated that the main photophysical properties of the dyes do not change if varying the labeling system from DNA to PNA. For both systems a fluorescence enhancement of the cyanine dyes bound to nucleic acids compared to free dyes is observed. Melting experiments were carried out to investigate the minimum length needed for efficient oligonucleotide-RNA binding. It was shown that shorter PNAs of 10 nts form a more stable duplex with RNA than the conventionally employed 18 nts long DNA oligonucleotides. Consequently, this allowed a reduction of the modification of the RNA, needed to bind the labels, towards a structure closer to the natural one. This was proven by the successful labeling of a new construct, D135-L1shL4 with an artificial loop L1sh shortened by 9 nts compared to D135-L14.

Additionally, an optimized annealing procedure for PNA-labels was developed, which leads to reduced dye and RNA degradation as well as RNA aggregation by avoiding the high temperatures needed for the annealing of DNA-labels. Furthermore, the labeling does not require an excess of PNA, in contrast to DNA. Together with the stronger binding affinity, this results in a higher labeling efficiency of PNA compared to DNA, observed in fluorescence native gels by a higher fraction of labeled molecules and less unbound labels. With regard to

FRET experiments this implies a decreased fluorescence background and a higher density of detectable molecules.

To confirm an unchanged activity of the RNA, cleavage assays were conducted. The activity of the ribozyme constructs (D135-L14, D135-L1shL4) was found to be similar when labeled with DNA or PNA.

The final interest for the PNA labels is their use for the study of RNA folding behavior by smFRET. Therefore, comparative smFRET experiments of the ribozyme model labeled with PNA or DNA dye conjugates were conducted. A similar folding pathway of the RNA for both labeling schemes was observed in the same FRET states 0.1, 0.2, 0.4 and 0.6 and overall similar trends of the state distributions. A more compact RNA structure is favored with the short PNAs in comparison to DNA, detected as shift to a higher FRET state. Nevertheless, the general folding behavior of the RNA is not changed by the PNA. Small variations are most likely due to the different structure of the loop carrying the label and from the folding heterogeneity of the construct. The dynamics of the RNA system also show similar trends in all labeling schemes. Notably, in all experiments, *i.e.* with DNA and with PNA labels, a previously unreported second family of transitions (0.1↔0.2↔0.5) was observed. These results indicate the existence of an additional alternative folding pathway.

A significant influence, *e.g.* on the free rotation of the dye, the labeling efficiency or the folding of the RNA by the presence or absence of a spacer group between the fluorophore attached to the PNA was not observed.

In summary, in this proof of concept study could be demonstrated that PNA oligonucleotides are perfect specific internal labeling agents for large RNA systems to study their folding mechanism without influencing the latter.

The remarkable potential of the PNA labeling has an effect in particular on RNA systems where modifications within the structure have a strong influence. For example, the folding of the *btuB* riboswitch of *E. coli* has been shown to be affected by internal sequence modifications and revealed very low labeling yields with DNA oligonucleotides. The above described labeling strategy was thus applied to the riboswitch. The labeling efficiency was remarkably higher when labeled with PNA instead of DNA. This gives the opportunity to investigate the riboswitch requiring reduced or even no modifications with the use of PNA labels.

Consequently, PNA can be employed in a fast, efficient and easy to handle labeling technique for various fluorescence applications, including smFRET for imaging of RNA folding processes *in vitr*o and in living cells.

Chapter 4 covers two collaborative projects. In another successful application of the PNA labeling scheme (*vide supra*), new PNA derivatives were synthesized to image the splicing reaction of a group II intron ribozyme by fluorescence techniques.

Finally a study is described which was undertaken with the aim to characterize the interaction of organometallic Rhenium-compounds with RNA by using the group II intron ribozyme construct D135-L14 as a model for large RNAs. It is shown that the Rhenium complexes, which are derivatized with a photo-labile protecting group and a nuclear localization sequence, can interact with RNA and reveal toxicity towards cervical cancer cells.

7 Zusammenfassung

Die Ribonucleinsäure (RNS) ist ein multifunktionelles Biomolekül, das für lebende Organismen essentiell ist, da es in unzähligen fundamentalen Prozessen involviert ist (Kapitel 1).

Der Fokus dieser Arbeit liegt auf der Untersuchung von Strategien zur Fluoreszenz-Markierung um komplexe Faltungsprozesse von RNS, im speziellen des *btuB* Riboschalter aus *Escherichia coli*, aufzuklären. Diese RNS besitzt eine genregulierende Funktion, die auf einer strukturellen Änderung aufgrund der spezifischen Bindung ihres Metaboliten Coenzym B_{12} (Adenosylcobalamin) basiert. Der Faltungsmechanismus dieser speziellen RNA und das Anbinden des Metaboliten sind noch nicht vollständig entschlüsselt und ihre Aufklärung könnte zusätzliche Details über die Funktionsweise von Riboschaltern und verwandter Systeme liefern. Aus diesem Wissen könnten sich viele Anwendungen ergeben, wie z.B. ihre Verwendung als hochgradig selektive antibakterielle Wirkstoffziele oder intrazelluläre Sensoren. Förster Resonanzenergietransfer (FRET) ist die Technik, die in den hier vorgestellten Studien zur Aufklärung der RNS-Faltung und der Metabolitenbindung angewendet wird. FRET ist ein distanzabhängiger Energietransfer von einem angeregten Donor-Fluorophor zu einem entsprechenden Akzeptor-Farbstoff. Mit dieser Technik können Abstandsänderungen zwischen den Fluorophoren nachverfolgt werden durch die Erfassung von Änderungen der Fluoreszenzintensitäten aufgrund des Energietransfers. Folglich kann man Informationen über Konformationen und Kinetiken von markierten Molekülen, die sich bewegen, erhalten. Diese Methode zeigt auf der Einzelmolekülebene besonders detaillierte Informationen über Faltungswege auf, die in Ensemblemessungen aufgrund der Mittelung einzelner Faltungsereignisse verborgen sind.

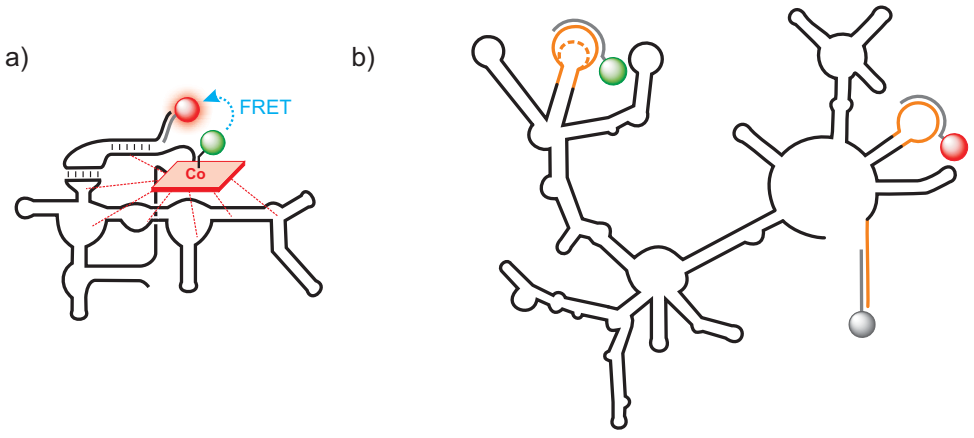


Abbildung 1: Neue Markierungsmethoden um die Faltung großer RNS mittels FRET zu untersuchen. a) Markierung des *btuB* Riboschalters und seines Metaboliten B_{12} mit Farbstoffen. b) Markierung eines Gruppe II Intron Ribozyms durch die Hybridisierung von fluoreszierenden PNS- oder DNS-Derivaten.

Für die Fluoreszenzdetektion müssen geeignete Marker am gefragten Molekül angebracht sein. Kapitel 2 setzt den Schwerpunkt auf die Farbstoffmarkierung des Metaboliten um sein Anbinden an die RNS und die Metabolit-induzierte Faltung der RNS zu untersuchen (Abbildung 1 a). Im Fall des *btuB* Riboschalters ist der Metabolit, Cobalamin (B_{12}), relativ groß. B_{12} selbst ist nicht fluoreszierend und kann deshalb mit Fluoreszenztechniken nicht sichtbar gemacht werden. Allerdings gibt seine Struktur die Möglichkeit den Metaboliten durch direktes Verknüpfen mit einem Farbstoffe fluoreszierend zu markieren. Für FRET wird ein geeignetes Farbstoffpaar benötigt, bei dem das Emissionsspektrum des Donor-Fluorophors und das Absorptionsspektrum des Akzeptors überlappen. Die fluoreszierenden Farbstoffe der Wahl sind Cy3 (Donor) und Cy5 (Akzeptor), ein bekanntes FRET-Paar, die für Untersuchungen von Nukleinsäuren etabliert sind. Alternativ kann Cy5 durch den nicht fluoreszierenden „Schwarzlochauslöscher" BHQ-2 (vom englischen *black hole quencher*) ersetzt werden. Die Verwendung von Fluoreszenzauslöschern minimiert mögliche Hintergrundprobleme ausgelöst durch die Fluoreszenz des markierten Metaboliten in hoher Konzentration. Der Donorfarbstoff kann an der RNS und der Akzeptorfarbstoff am B_{12} angebracht sein, oder umgekehrt. Die axiale Position wurde zur Anbringung des Farbstoffes ausgewählt, da es sich gezeigt hat, dass diese Position des Cobalamins zwar die Bindungsstärke des B_{12} zum Riboschalter beeinflusst, nicht aber das Faltungsverhalten der RNA verändert. Modifikationen am Corrinring hingegen würden zu einer veränderten Faltung der RNS führen. Deshalb wurde eine verknüpfende Gruppe am Kobaltzentrum auf der β-Seite des B_{12} angebracht, an die die Farbstoffe mittels NHS-Ester-Chemie gekoppelt wurden.

Eine wichtige Annahme bei der Verwendung dieser Methodik ist, dass die modifizierten B_{12}-Derivate die Faltung des Riboschalters nicht drastisch verändern. Daher wurden zusätzlich zur Synthese und Charakterisierung der neuen farbstoffmarkierten B_{12}-Derivate, deren Wechselwirkungen mit dem *btuB* Riboschalter mittels „in-line probing" Experimenten im Vergleich zum natürlichen Liganden Adenosylcobalamin untersucht. Es wurde herausgefunden, dass mit sulfonierten, und damit negativ geladenen, Cyaninfarbstoffen markierte B_{12}-Derivate keine geeigneten Proben für FRET Experimente mit dem *btuB* Riboschalter sind. **B_{12}-Sulfo-Cy3** verändert das Faltungsverhalten drastisch und **B_{12}-Sulfo-Cy5** bindet überhaupt nicht an den Riboschalter. Diese Resultate unterstreichen die Spezifität des Riboschalters gegenüber unterschiedlichen B_{12}-Derivaten. Die Bindungsaffinitäten von **B_{12}-Cy3**, **B_{12}-Cy5** und **B_{12}-BHQ-2** sind, wie erwartet, reduziert verglichen mit Adenosylcobalamin, aber sind im Bereich von Vitamin B_{12} (~ 200 µM). Desweiteren wurde beobachtet, dass das Anbringen der sperrigen Farbstoffe (Cy3, Cy5 oder BHQ-2) an das B_{12} die konformelle RNS-Schaltung geringfügig beeinflusst. Allerdings sind

die Änderungen der Riboschalterfaltung ziemlich gering. **B_{12}-Cy3**, **B_{12}-Cy5** und **B_{12}-BHQ-2** sind deshalb vielversprechende Kandidaten für FRET Experimente.
Fluoreszenzexperimente haben gezeigt, dass bei der Wahl der Messbedingungen für FRET Experimente die potentielle Photolyse der Co-C-Bindung berücksichtigt werden muss, da sie zu einer Abspaltung des Fluorophors führen kann. Zudem wurde beobachtet, dass B_{12} selbst die Fluoreszenz des verknüpften Farbstoffs bis zu einem gewissen Grad auslöschen kann.
Nichtsdestotrotz bietet diese neue Strategie der Fluoreszenzmarkierung des Metaboliten die Möglichkeit Einblicke in dessen Bindungsmechanismus zur RNS zu erhalten, die sonst nicht zu erreichen wären. Darüber hinaus haben Untersuchungen der modifizierten B_{12}-Derivate mit dem *btuB* Riboschalter gezeigt, dass negative Ladungen am B_{12} die Faltung des Riboschalters beeinflussen.

Kapitel 3 behandelt die Fluoreszenzmarkierung der RNS (Abbildung 1 b). Positionsspezifische interne Markierungsmethoden für große RNS gibt es zwar, aber sie leiden unter unzulänglichen Markierungsausbeuten und einer aufwendigen Synthese. Übliche indirekte Methoden wiederum benötigen relativ große Modifikationen der RNS, an die DNS-Farbstoff-Konjugate hybridisiert werden können, was eine Fehlfaltung der RNS verursachen kann. Basierend auf diesem bekannten Schema haben wir die Markierungsstrategie durch die Verwendung von Peptidnukleinsäuren (PNS), anstelle von DNS, verbessert. PNS ist ein ungeladenes, künstliches Nukleinsäureanaloga, welches bekannt dafür ist stärker an die RNS zu binden als es DNS kann. Folglich können die Modifikationen der RNS minimiert werden um den Marker zu binden.
Neue PNS-Derivate, die sulfonierte Cy3 oder Cy5 FRET Farbstoffe tragen, entweder direkt oder über einen zusätzlichen Abstandshalters angebunden, wurden mittels Festphasensynthese synthetisiert und dann charakterisiert. Die Farbstoffkonjugate wurden an einem genau untersuchten Modellsystem getestet, was einen direkten Vergleich des zuvor verwendeten Markierungsschemas mit dem Verbesserten ermöglicht. Hierfür haben wir eine verkürzte Version des mitochondrialen Gruppe II Introns *Sc*.ai5γ aus *Saccharomyces cerevisiae* (D135-L14, Abbildung 1 b) ausgewählt. Es ist ein katalytisch aktives großes RNS Molekül (> 600 Nukleotide), das bereits erfolgreich mittels smFRET untersucht wurde. Daher würden etwaige Beeinträchtigungen der RNS-Faltung durch die PNA-Markierung leicht erkannt werden.
Die Fluoreszenzeigenschaften der Proben sind unter den Einflüssen, die für eine optimale Fluoreszenzmarkierung berücksichtigt werden müssen. Ein wichtiger Aspekt ist ihre Abhängigkeit von Umgebungseinflüssen. Fluoreszenzquantenausbeuten und statische Anisotropiemessungen haben gezeigt, dass sich die wichtigsten photophysikalischen Eigenschaften der Farbstoffe nicht verändern wenn man das Markierungssystem von DNS zu PNS ändert. Bei beiden Systemen wird eine Fluoreszenzverstärkung der Cyaninfarbstoffe

durch das Anbinden an die Nukleinsäuren im Vergleich zu den freien Farbstoffen beobachtet. Schmelzexperimente wurden durchgeführt um die mindestens benötigte Länge für eine effiziente Oligonukleotid-RNS Bindung zu untersuchen. Es hat sich gezeigt, dass kürzere PNS aus 10 Nukleotiden eine stabilere Duplex mit RNS bilden als die üblicherweise verwendeten 18 Nukleotide langen DNS Oligonukleotide. Folglich ermöglicht dies eine Verringerung der Modifikation der RNS, die für die Anbringung der Markierung benötigt wird, in Richtung einer Struktur, die eher der natürlichen entspricht. Dies wurde mit der erfolgreichen Markierung eines neuen Konstrukts, D135-L1shL4 mit einer künstlichen Schlaufe L1sh, die im Vergleich zu D135-L14 um 9 Nukleotide gekürzt ist, bewiesen.

Zusätzlich wurde eine optimierte Anbindungsprozedur für die PNS-Marker entwickelt, welche zu reduzierter Farbstoff- und RNS-Degradation sowie RNS-Aggregation führt, indem die hohen Temperaturen, die zur Anbindung der DNS-Marker benötigt werden, vermieden werden. Desweiteren. wird kein Überschuss von PNS benötig, im Gegensatz zur DNS. Unter gemeinsamer Berücksichtigung der stärkeren Bindungsaffinität führt dies zu einer höheren Markierungseffizienz von PNS gegenüber DNS, was in nativen Fluoreszenzgelen anhand einer größeren Fraktion an markierten Molekülen und weniger ungebunden Markern beobachtet wurde. Im Hinblick auf FRET Experimente impliziert dies einen reduzierten Fluoreszenzhintergrund und eine höhere Dichte an detektierbaren Molekülen.

Um eine unveränderte Aktivität der RNS zu bestätigen wurden Spaltungs-Assays durchgeführt. Es wurden gleiche Aktivitäten der Ribozymkonstrukte (D135-L14, D135-L1shL4) bei Markierung mit DNS oder PNS beobachtet.

Das schlussendliche Interesse an den PNS-Markern ist ihre Verwendung zur Untersuchung des Faltungsverhaltens von RNS mittels smFRET. Daher wurden Vergleichsexperimente am Ribozymmodell, das mit PNS- oder DNS-Farbstoff-Konjugaten markiert wurde, durchgeführt. Es wurde der gleiche Faltungsweg der RNS für beide Markierungsschemata anhand der gleichen FRET Zustände 0.1, 0.2, 0.4 und 0.6 und im Großen und Ganzen gleichen Trends in den Zustandsverteilungen beobachtet. Eine kompaktere RNS-Struktur ist mit den kurzen PNS begünstigt verglichen mit DNS, was sich anhand einer Verschiebung zu einem höheren FRET Zustand bemerkbar macht. Nichtsdestotrotz wird das generelle Faltungsverhalten der RNS nicht durch die PNS verändert. Kleine Variationen kommen wahrscheinlich von einer veränderten Struktur der Schlaufe, die den Marker trägt, und von der Faltungsheterogenität des Konstrukts. Die Dynamiken des RNS Systems zeigen zudem gleiche Trends mit allen Markierungsschemata. Auffälliger Weise wurde in allen Experimenten, also mit DNS- und mit PNS-Markern, eine zweite Familie von Übergängen beobachtet (0.1↔0.2↔0.5), die bisher noch nicht beschrieben wurde. Diese Ergebnisse lassen vermuten, dass ein zusätzlicher alternativer Faltungsweg existiert.

Einen signifikanten Einfluss durch die An-oder Abwesenheit einer Abstandshaltergruppe zwischen der PNS und dem Farbstoff, z.B. auf die freie Drehbarkeit des Farbstoffs, die Markierungseffizienz oder die Faltung der RNS hat man nicht festgestellt.
Zusammenfassend konnte in dieser Machbarkeitsstudie gezeigt werden, dass PNS Oligonukleotide perfekte spezifische interne Markierungsstoffe für große RNS Systeme darstellen zur Untersuchung ihres Faltungsmechanismus ohne letzteren zu beeinflussen.
Das bemerkenswerte Potential von PNS kommt besonders dort zum Tragen wo Modifikationen innerhalb der Struktur einen großen Einfluss ausüben. Zum Beispiel hat sich für die Faltung des *btuB* Riboschalters aus *E. coli* eine Beeinflussung durch interne Sequenzmodifikationen und eine sehr geringe Markierungsausbeute mit DNS-Oligonukleotiden gezeigt. Die oben beschriebene Markierungsmethode wurde deshalb auf den Riboschalter angewendet. Die Markierungseffizienz war bei Markierung mit PNS statt mit DNS beträchtlich höher. Dies gibt die Möglichkeit den Riboschalter mit verringerten oder sogar ohne Modifizierungen durch Verwendung von PNS-Markern zu untersuchen.
Somit kann PNS als schnelle und effiziente Markierungstechnik mit bequemer Handhabung für verschiedenste Fluoreszenzanwendung verwendet werden, eingeschlossen smFRET zur Abbildung von RNS Faltungsprozessen *in vitro* und in lebenden Zellen.

Kapitel 4 behandelt zwei gemeinschaftliche Projekte. In einer weiteren erfolgreichen Anwendung des PNS Markierungsschemas (siehe oben) wurden neue PNS Derivate synthetisiert um die Splicing-Reaktion eines Gruppe II Intron Ribozyms zu untersuchen.
Zum Abschluss wird eine Studie beschrieben, die mit dem Ziel durchgeführt wurde, die Wechselwirkung von organometallischen Rhenium-Verbindungen mit RNS unter Verwendung des Gruppe II Intron Ribozymkonstrukts D135-L14 als Modellsystem für große RNS zu charakterisieren. Es wird gezeigt, dass die Rhenium-Komplexe, die mit einer photo-labilen Schutzgruppe und einer Kernlokalisierungssequenz derivatisiert sind, mit RNS wechselwirken können und toxisch gegenüber zervikalen Krebszellen sind.

8 Literature

[1] *The American Heritage dictionary of the English language*, Houghton Mifflin Company, **2000**.

[2] F. Crick, *Nature* **1970**, *227*, 561–563.

[3] K. Hoogsteen, *Acta Crystallogr.* **1963**, *16*, 907–916.

[4] E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei, H. M. Al-Hashimi, *Nature* **2011**, *470*, 498–502.

[5] F. H. C. Crick, *J. Mol. Biol.* **1966**, *19*, 548–555.

[6] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.

[7] L. Pauling, R. B. Corey, *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 84–97.

[8] L. Pauling, R. B. Corey, *Nature* **1953**, *171*, 346.

[9] D. W. Staple, S. E. Butcher, *PLoS Biol.* **2005**, *3*, e213.

[10] S. D. Copley, E. Smith, H. J. Morowitz, *Bioorg. Chem.* **2007**, *35*, 430–443.

[11] G. F. Joyce, *Nature* **2002**, *418*, 214–221.

[12] R. Das, L. W. Kwok, I. S. Millett, Y. Bai, T. T. Mills, J. Jacob, G. S. Maskel, S. Seifert, Mochrie, Simon G. J., P. Thiyagarajan et al., *J. Mol. Biol.* **2003**, *332*, 311–319.

[13] S. A. Woodson, *Curr. Opin. Chem. Biol.* **2005**, *9*, 104–109.

[14] R. K. O. Sigel, A. Vaidya, A. M. Pyle, *Nat. Struct. Biol.* **2000**, *7*, 1111–1116.

[15] K. F. Blount, R. R. Breaker, *Nat. Biotechnol.* **2006**, *24*, 1558–1564.

[16] T. Förster, *Z. Naturforsch.* **1949**, *4a*, 321–327.

[17] J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Springer, New York, **2006**.

[18] R. Roy, S. Hohng, T. Ha, *Nat. Methods* **2008**, *5*, 507–516.

[19] J. M. Dixon, M. Taniguchi, J. S. Lindsey, *Photochem. Photobiol.* **2005**, 212–213.

[20] B. J. Harvey, C. Perez, M. Levitus, *Photochem. Photobiol. Sci.* **2009**, *8*, 1105–1110.

[21] M. E. Sanborn, B. K. Connolly, K. Gurunathan, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 11064–11074.

[22] K. Jia, Y. Wan, A. Xia, S. Li, F. Gong, G. Yang, *J. Phys. Chem. A* **2007**, *111*, 1593–1597.

[23] R. B. Mujumdar, L. A. Ernst, S. R. Mujumdar, C. J. Lewis, A. S. Waggoner, *Bionconj. Chem.* **1993**, 105–111.

[24] L. Le Reste, J. Hohlbein, K. Gryte, A. N. Kapanidis, *Biophys. J.* **2012**, *102*, 2658–2668.

[25] M. Ueda, Y. Sako, T. Tanaka, P. Devreotes, T. Yanagida, *Science* **2001**, *294*, 864–867.

[26] A. Moter, U. B. Göbel, *J. Microbiol. Methods* **2000**, *41*, 85–112.

[27] M. V. Kvach, A. V. Ustinov, I. A. Stepanova, A. D. Malakhov, M. V. Skorobogatyi, V. V. Shmanai, V. A. Korshun, *Eur. J. Org. Chem.* **2008**, *2008*, 2107–2117.

[28] B. J. Harvey, M. Levitus, *J. Fluoresc.* **2009**, *19*, 443–448.

[29] J. Widengren, P. Schwille, *J. Phys. Chem. A* **2000**, *104*, 6416–6428.

[30] L. Urnavicius, S. A. McPhee, D. M. J. Lilley, D. G. Norman, *Biophys. J.* **2012**, *102*, 561–568.

[31] A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 11176–11181.

[32] A. Iqbal, L. Wang, K. C. Thompson, D. M. J. Lilley, D. G. Norman, *Biochemistry* **2008**, *47*, 7857–7862.

[33] J. Spiriti, J. K. Binder, M. Levitus, A. van der Vaart, *Biophys. J.* **2011**, *100*, 1049–1057.

[34] J. Ouellet, S. Schorr, A. Iqbal, T. J. Wilson, D. M. Lilley, *Biophys. J.* **2011**, *101*, 1148–1154.

[35] S. Weiss, *Nat. Struct. Biol.* **2000**, *7*, 724–729.

[36] C. E. Aitken, J. D. Puglisi, *Nat. Struct. Mol. Biol.* **2010**, *17*, 793–800.

[37] A. N. Kapanidis, S. Weiss, *J. Chem. Phys.* **2002**, *117*, 10953–10964.

[38] S. Weiss, *Science* **1999**, *283*, 1676–1683.

[39] J. Chen, R. V. Dalal, A. N. Petrov, A. Tsai, S. E. O'Leary, K. Chapin, J. Cheng, M. Ewan, P.-L. Hsiung, P. Lundquist et al., *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 664–669.

[40] A. Haller, R. B. Altman, M. F. Souliere, S. C. Blanchard, R. Micura, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4188–4193.

[41] R. Zhao, D. Rueda, *Methods* **2009**, *49*, 112–117.

[42] M. Steiner, K. S. Karunatilaka, R. K. O. Sigel, D. Rueda, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13853–13858.

[43] M. Blanco, N. G. Walter, *Methods Enzymol.* **2010**, 153–178.

[44] T. Ha, *Methods* **2001**, *25*, 78–86.

[45] I. Rasnik, S. A. McKinney, T. Ha, *Nat. Methods* **2006**, *3*, 891–893.

[46] A. Nahvi, J. E. Barrick, R. R. Breaker, *Nucleic Acids Res.* **2004**, *32*, 143–150.

[47] C. C. Fowler, E. D. Brown, Y. Li, *Chem. Biol.* **2010**, *17*, 756–765.

[48] L. Martini, S. S. Mansy, *Chem. Commun.* **2011**, *47*, 10734–10736.

[49] E. E. Regulski, R. R. Breaker, *Methods Mol. Biol.* **2008**, *419*, 53–67.

[50] A. Savinov, C. F. Perez, S. M. Block, *Biochim. Biophys. Acta, Gene Regul. Mech.* **2014**, *1839*, 1030–1045.

[51] D. Crowfoot-Hodgkin, *Angew. Chem.* **1965**, *77*, 954–962.

[52] D. C. Hodgkin, J. Kamper, J. Lindsey, M. MacKay, J. Pickworth, J. H. Robertson, C. B. Shoemaker, J. G. White, R. J. Prosen, K. N. Trueblood, *Proc. R. Soc. London, Ser. A* **1957**, *242*, 228–263.

[53] B. Kräutler, D. Arigoni, B. T. Golding, *Vitamin B_{12} and B_{12}-proteins*, Wiley-VCH, Weinheim, **1998**.

[54] S. Gschösser, K. Gruber, C. Kratky, C. Eichmüller, B. Kräutler, *Angew. Chem. Int. Ed.* **2005**, *44*, 2284–2288.

[55] W. Fieber, B. Hoffmann, W. Schmidt, E. Stupperich, R. Konrat, B. Kräutler, *Helv. Chim. Acta* **2002**, *85*, 927–944.

[56] K. Gruber, B. Puffer, B. Kräutler, *Chem. Soc. Rev.* **2011**, *40*, 4346–4363.

[57] H. A. Barker, R. D. Smyth, H. Weissbach, J. I. Toohey, J. N. Ladd, B. E. Volcani, *J. Biol. Chem.* **1960**, *235*, 480–488.

[58] P. G. Lenhert, D. C. Hodgkin, *Nature* **1961**, *192*, 937–938.

[59] B. D. Martin, R. G. Finke, *J. Am. Chem. Soc.* **1992**, *114*, 585–592.

[60] B. D. Martin, R. G. Finke, *J. Am. Chem. Soc.* **1990**, *112*, 2419–2420.

[61] J. Halpern, *Science* **1985**, *227*, 869–875.

[62] T. A. Shell, D. S. Lawrence, *J. Am. Chem. Soc.* **2011**, *133*, 2148–2150.

[63] B. Kräutler, W. Keller, C. Kratky, *J. Am. Chem. Soc.* **1989**, *111*, 8936–8938.

[64] G. N. Schrauzer, L.-P. Lee, J. W. Sibert, *J. Am. Chem. Soc.* **1970**, *92*, 2997–3005.

[65] F. Mancia, N. H. Keep, A. Nakagawa, P. F. Leadlay, S. McSweeney, B. Rasmussen, P. Bösecke, O. Diat, P. R. Evans, *Structure* **1996**, *4*, 339–350.

[66] R. Reitzer, K. Gruber, G. Jogl, U. G. Wagner, H. Bothe, W. Buckel, C. Kratky, *Structure* **1999**, *7*, 891–902.

[67] L. Randaccio, S. Geremia, N. Demitri, J. Wuerges, *Molecules* **2010**, *15*, 3228–3259.

[68] D. P. Chimento, A. K. Mohanty, R. J. Kadner, M. C. Wiener, *Nat. Struct. Biol.* **2003**, *10*, 394–401.

[69] D. D. Shultis, M. D. Purdy, C. N. Banchs, M. C. Wiener, *Science* **2006**, *312*, 1396–1399.

[70] K. P. Locher, A. T. Lee, D. C. Rees, *Science* **2002**, *296*, 1091–1098.

[71] N. K. Karpowich, H. H. Huang, P. C. Smith, J. F. Hunt, *J. Biol. Chem.* **2003**, *278*, 8429–8434.

[72] R. N. Hvorup, B. A. Goetz, M. Niederer, K. Hollenstein, E. Perozo, K. P. Locher, *Science* **2007**, *317*, 1387–1390.

[73] A. Roth, R. R. Breaker, *Annu. Rev. Biochem.* **2009**, *78*, 305–334.

[74] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043–1049.

[75] A. Roth, W. C. Winkler, E. E. Regulski, B. W. K. Lee, J. Lim, I. Jona, J. E. Barrick, A. Ritwik, J. N. Kim, R. Welz et al., *Nat. Struct. Mol. Biol.* **2007**, *14*, 308–317.

[76] M. M. Meyer, A. Roth, S. M. Chervin, G. A. Garcia, R. R. Breaker, *RNA* **2008**, *14*, 685–695.

[77] I. Gusarov, E. Nudler, *Mol. Cell* **1999**, *3*, 495–504.

[78] W. S. Yarnell, J. W. Roberts, *Science* **1999**, *284*, 611–615.

[79] J. Shine, L. Dalgarno, *Proc. Natl. Acad. Sci. USA* **1974**, *71*, 1342–1346.

[80] R. T. Fuchs, F. J. Grundy, T. M. Henkin, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 4876–4880.

[81] M. F. Soulière, A. Haller, T. Santner, R. Micura, *Angew. Chem. Int. Ed.* **2013**, *52*, 1874–1877.

[82] A. Peselis, A. Serganov, *Nat. Struct. Biol.* **2012**, *19*, 1182–1184.

[83] J. E. Johnson Jr., F. E. Reyes, J. T. Polaski, R. T. Batey, *Nature* **2012**, *492*, 133–137.

[84] X. Nou, R. J. Kadner, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 7190–7195.

[85] G. A. Perdrizet, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 3323–3328.

[86] S. Gallo, *Dissertation*, Universität Zürich, Zürich, **2009**.

[87] S. Gallo, M. Oberhuber, R. K. O. Sigel, B. Kräutler, *ChemBioChem* **2008**, *9*, 1408–1414.

[88] S. Gallo, S. Mundwiler, R. Alberto, R. K. O. Sigel, *Chem. Commun.* **2010**, *47*, 403–405.

[89] M. Oivanen, S. Kuusela, H. Lönnberg, *Chem. Rev.* **1998**, *98*, 961–990.

[90] A. Haller, U. Rieder, M. Aigner, S. C. Blanchard, R. Micura, *Nat. Chem. Biol.* **2011**, *7*, 393–400.

[91] Synthesis strategy and precursors N-((1E)-3-(phenylimino)prop-1-enyl)benzenamine hydrochloride and 1-(5-carboxypentyl)-2,3,3-trimethyl-3H-indolium chloride were kindly provided by Dr. Miquel Barceló Oliver.

[92] Sonja Giger and Adrian Kämpfer are acknowledged for their contribution to the synthesis.

[93] M. E. Jung, W.-J. Kim, *Bioorg. Med. Chem.* **2006**, *14*, 92–97.

[94] C. C. Smeltzer, M. J. Cannon, P. R. Pinson, J. D. Munger, F. G. West, C. B. Grissom, *Org. Lett.* **2001**, *3*, 799–801.

[95] B. Kräutler, *Met. Ions Life Sci.* **2009**, *6*, 1–51.

[96] M. F. Summers, L. G. Marzilli, A. Bax, *J. Am. Chem. Soc.* **1986**, *108*, 4285–4294.

[97] R. K. O. Sigel, E. Freisinger, B. Lippert, *J. Biol. Inorg. Chem.* **2000**, *5*, 287–299.

[98] M. Lee, C. B. Grissom, *Org. Lett.* **2009**, *11*, 2499–2502.

[99] A. H. Gore, M. B. Kale, P. V. Anbhule, S. R. Patil, G. B. Kolekar, *RSC Adv.* **2014**, *4*, 683–692.

[100] E. Vaishnavi, R. Renganathan, *Spectrochim. Acta, Part A* **2013**, *115*, 603–609.

[101] D. G. Norman, R. J. Grainger, D. Uhrín, D. M. J. Lilley, *Biochemistry* **2000**, *39*, 6317–6324.

[102] P. Zhu, H. G. Craighead, *Annu. Rev. Biophys.* **2012**, *41*, 269–293.

[103] B. Okumus, T. J. Wilson, D. M. J. Lilley, T. Ha, *Biophys. J.* **2004**, *87*, 2798–2806.

[104] S. Gallo, E. Freisinger, R. K. O. Sigel, *Inorg. Chim. Acta* **2007**, *360*, 360–368.

[105] Q. Meng, M. Yu, H. Zhang, J. Ren, D. Huang, *Dyes Pigm.* **2007**, *73*, 254–260.

[106] G. J. Smith, T. R. Sosnick, N. F. Scherer, T. Pan, *RNA* **2005**, *11*, 234–239.

[107] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.

[108] T. R. Cech, *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 759–768.

[109] S. Altman, *Angew. Chem. Int. Ed. Engl.* **1990**, *29*, 749–758.

[110] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, G. Bruening, *Science* **1986**, *231*, 1577–1580.

[111] J. M. Buzayan, W. L. Gerlach, G. Bruening, *Nature* **1986**, *323*, 349–353.

[112] L. Sharmeen, M. Y. Kuo, G. Dinter-Gottlieb, J. Taylor, *J. Virol.* **1988**, *62*, 2674–2679.

[113] S. E. Butcher, *Curr. Opin. Chem. Biol.* **2001**, *11*, 315–320.

[114] T. Yoshihisa, *Front. Genet.* **2014**, *5*, 213.

[115] N. G. Walter, D. R. Engelke, *The Biologist* **2002**, *49*, 199–203.

[116] D. M. J. Lilley, *Philos. Trans. R. Soc., B* **2011**, *366*, 2910–2917.

[117] H. D. Robertson, S. Altman, J. D. Smith, *J. Biol. Chem.* **1972**, *247*, 5243–5251.

[118] M. F. Baer, R. M. Reilly, G. M. McCorkle, T. Y. Hai, S. Altman, U. L. RajBhandary, *J. Biol. Chem.* **1988**, *263*, 2344–2351.

[119] D. M. J. Lilley, *Curr. Opin. Chem. Biol.* **2005**, *15*, 313–323.

[120] M. Steiner, *Dissertation*, Universität Zürich, Zürich, **2008**.

[121] K. A. Jarrell, C. L. Peebles, R. C. Dietrich, S. L. Romiti, P. S. Perlman, *J. Biol. Chem.* **1988**, *263*, 3432–3439.

[122] C. L. Peebles, P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell, H.-L. Cheng, *Cell*, *44*, 213–223.

[123] M. Podar, P. S. Perlman, R. A. Padgett, *Mol. Cell. Biol.* **1995**, *15*, 4466–4478.

[124] O. Fedorova, N. Zingler, *Biol. Chem.* **2007**, *388*, 665–678.

[125] F. Michel, J. Ferat, *Annu. Rev. Biochem.* **1995**, *64*, 435–461.

[126] N. Toor, A. R. Robart, J. Christianson, S. Zimmerly, *Nucleic Acids Res.* **2006**, *34*, 6461–6471.

[127] P. Z. Qin, A. M. Pyle, *Curr. Opin. Chem. Biol.* **1998**, *8*, 301–308.

[128] P. Z. Qin, A. M. Pyle, *Biochemistry* **1997**, *36*, 4718–4730.

[129] A. Jacquier, F. Michel, *Cell* **1987**, *50*, 17–29.

[130] U. Schmidt, M. Podar, U. Stahl, P. S. Perlman, *RNA* **1996**, *2*, 1161–1172.

[131] J. L. Koch, S. C. Boulanger, S. D. Dib-Hajj, S. K. Hebbar, P. S. Perlman, *Mol. Cell. Biol.* **1992**, *12*, 1950–1958.

[132] O. Fedorova, T. Mitros, A. M. Pyle, *J. Mol. Biol.* **2003**, *330*, 197–209.

[133] H. Wank, J. SanFilippo, R. N. Singh, M. Matsuura, A. M. Lambowitz, *Mol. Cell* **1999**, *4*, 239–250.

[134] A. Jacquier, F. Michel, *J. Mol. Biol.* **1990**, *213*, 437–447.

[135] M. Boudvillain, A. de Lencastre, A. M. Pyle, *Nature* **2000**, *406*, 315–318.

[136] C. L. Harris-Kerr, M. Zhang, C. L. Peebles, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 10658–10662.

[137] R. K. O. Sigel, D. G. Sashital, D. L. Abramovitz, A. G. Palmer, S. E. Butcher, A. M. Pyle, *Nat. Struct. Mol. Biol.* **2004**, *11*, 187–192.

[138] J. F. Swisher, C. M.Duarte, L. J. Su, A. M. Pyle, *EMBO J.* **2001**, 2051–2061.

[139] D. Kruschel, M. Skilandat, R. K. O. Sigel, *RNA* **2014**, *20*, 295–307.

[140] M. Skilandat, R. K. O. Sigel, *J. Biol. Chem.* **2014**, *112*, 3403–3408.

[141] M. Pechlaner, R. K. O. Sigel, W. F. van Gunsteren, J. Dolenc, *Biochemistry* **2013**, *52*, 7099–7113.

[142] N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, *Science* **2008**, *320*, 77–82.

[143] O. Fedorova, C. Waldsich, A. M. Pyle, *J. Mol. Biol.* **2007**, *366*, 1099–1114.

[144] J. F. Swisher, L. J. Su, M. Brenowitz, V. E. Anderson, A. M. Pyle, *J. Mol. Biol.* **2002**, *315*, 297–310.

[145] L. J. Su, M. Brenowitz, A. M. Pyle, *J. Mol. Biol.* **2003**, *334*, 639–652.

[146] E. Paredes, M. Evans, S. R. Das, *Methods* **2011**, *54*, 251–259.

[147] P. Z. Qin, A. M. Pyle, *Methods* **1999**, *18*, 60–70.

[148] K. Onizuka, A. Shibata, Y. Taniguchi, S. Sasaki, *Chem. Commun.* **2011**, *47*, 5004–5006.

[149] L. Büttner, F. Javadi-Zarnaghi, C. Höbartner, *J. Am. Chem. Soc.* **2014**, *136*, 8131–8137.

[150] M. J. Moore, P. A. Sharp, *Science* **1992**, *256*, 992–997.

[151] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, *Science* **1991**, *254*, 1497–1500.

[152] P. E. Nielsen, *Peptide nucleic acids*, Horizon Bioscience, Wymondham, **2004**.

[153] K. K. Jensen, H. Ørum, P. E. Nielsen, B. Nordén, *Biochemistry* **1997**, *36*, 5072–5077.

[154] M. Egholm, O. Buchardt, L. Christensen, C. Behrens, S. M. Freier, D. A. Driver, R. H. Berg, S. K. Kim, B. Norden, P. E. Nielsen, *Nature* **1993**, *365*, 566–568.

[155] A. M. Blanco, R. Artero, *Methods* **2010**, *52*, 343–351.

[156] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154.

[157] L. A. Carpino, *J. Am. Chem. Soc.* **1993**, *115*, 4397–4398.

[158] E. Kaiser, R. L. Colescott, C. D. Bossinger, P. I. Cook, *Anal. Biochem.* **1970**, *34*, 595–598.

[159] V. K. Sarin, S. B. H. Kent, J. P. Tam, R. B. Merrifield, *Anal. Biochem.* **1981**, *117*, 147–157.

[160] J. C. Norton, M. A. Piatyszek, W. E. Wright, J. W. Shay, D. R. Corey, *Nat. Biotechnol.* **1996**, *14*, 615–619.

[161] U. Koppelhus, V. Zachar, P. E. Nielsen, X. Liu, J. Eugen-Olsen, P. Ebbesen, *Nucleic Acids Res.* **1997**, *25*, 2167–2173.

[162] R. W. Taylor, P. F. Chinnery, D. M. Turnbull, R. N. Lightowlers, *Nat. Genet.* **1997**, *15*, 212–215.

[163] D. A. Braasch, D. R. Corey, *Biochemistry* **2002**, *41*, 4503–4510.

[164] P. E. Nielsen, *Curr. Opin. Biotechnol.* **2001**, *12*, 16–20.

[165] J. O. Smith, D. A. Olson, B. A. Armitage, *J. Am. Chem. Soc.* **1999**, *121*, 2686–2695.

[166] K. K. Sadhu, N. Winssinger, *Chem. Eur. J.* **2013**, *19*, 8182–8189.

[167] A. G. Torres, M. M. Fabani, E. Vigorito, D. Williams, N. Al-Obaidi, F. Wojciechowski, R. H. E. Hudson, O. Seitz, M. J. Gait, *Nucleic Acids Res.* **2012**, *40*, 2152–2167.

[168] K. L. Robertson, L. Yu, B. A. Armitage, A. J. Lopez, L. A. Peteanu, *Biochemistry* **2006**, *45*, 6066–6074.

[169] A. G. Schmitz, S. Zelger-Paulus, G. Gasser, R. K. O. Sigel, *ChemBioChem* **2015**, accepted.

[170] P. Anstaett, Y. Zheng, T. Thai, A. M. Funston, U. Bach, G. Gasser, *Angew. Chem. Int. Ed.* **2013**, *52*, 4217–4220.

[171] L. D. Mayfield, D. R. Corey, *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1419–1422.

[172] A. M. Sosniak, G. Gasser, N. Metzler-Nolte, *Org. Biomol. Chem.* **2009**, *7*, 4992.

[173] http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi.

[174] S. A. McKinney, C. Joo, T. Ha, *Biophys. J.* **2006**, *91*, 1941–1951.

[175] S. L. B. König, M. Hadzic, E. Fiorini, R. Börner, D. Kowerko, W. U. Blanckenhorn, R. K. O. Sigel, *PLoS ONE* **2013**, *8*, e84157.

[176] P. Choudhary, *Dissertation*, Universität Zürich, Zürich, **2013**.

[177] A. Aletras, K. Barlos, D. Gatos, S. Koutsogianni, P. Mamos, *Int. J. Pept. Protein Res.* **1995**, *45*, 488–496.

[178] A. Leonidova, V. Pierroz, R. Rubbiani, Y. Lan, A. G. Schmitz, A. Kaech, R. K. O. Sigel, S. Ferrari, G. Gasser, *Chem. Sci.* **2014**, *5*, 4044–4056.

[179] M. Hadzic, D. Kowerko, R. Börner, S. L. B. König, R. K. O. Sigel, *manuscript in preparation*.

[180] S. C. Walker, J. M. Avis, G. L. Conn, *Nucleic Acids Res.* **2003**, *31*, e82.

[181] S. Gallo, M. Furler, R. K. O. Sigel, *Chimia* **2005**, *59*, 812–816.

[182] Susann Zelger-Paulus is acknowledged for the PCR of D135-L14sh and D135-L1shL4.

[183] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, S. Weiss, *Acc. Chem. Res.* **2005**, *38*, 523–533.

9 Appendix

A. MALDI-MS Spectra and HPLC Chromatograms

B. Melting Curves

C. Single-Turnover Cleavage Assays

D. Plasmid Table pSPA7a

E. UV/Vis Spectra

F. HR-ESI-MS Spectra

G. NMR Spectra

H. Time Trajectories of smFRET Experiments

I. *btuB* Constructs

A. MALDI-MS Spectra and HPLC Chromatograms

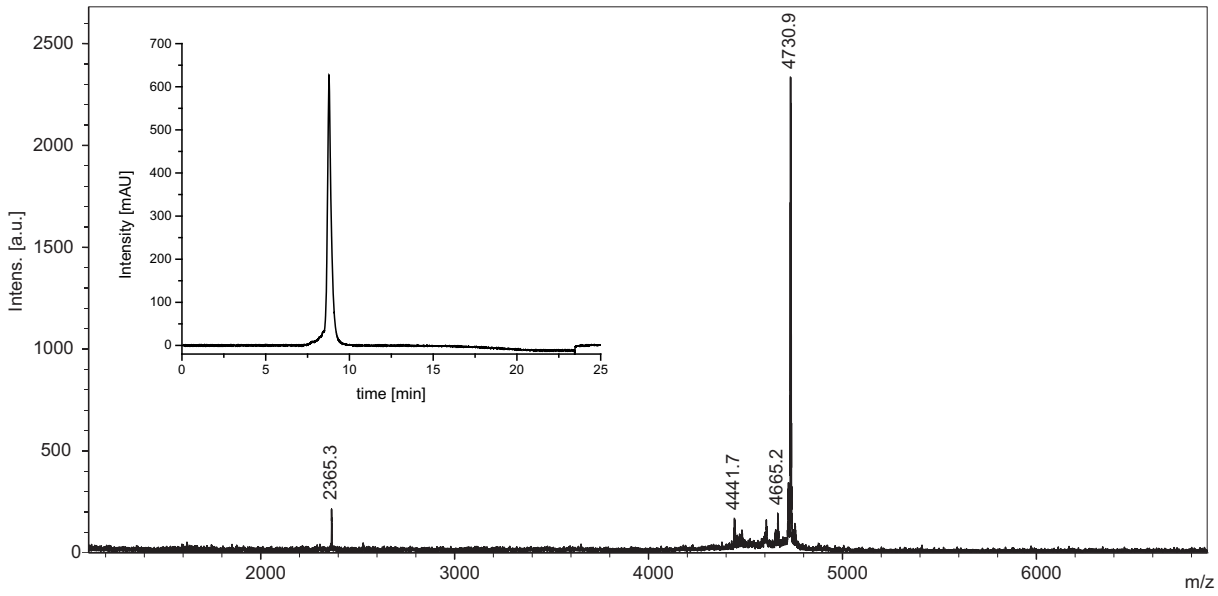


Figure A1: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA1S-Cy5.

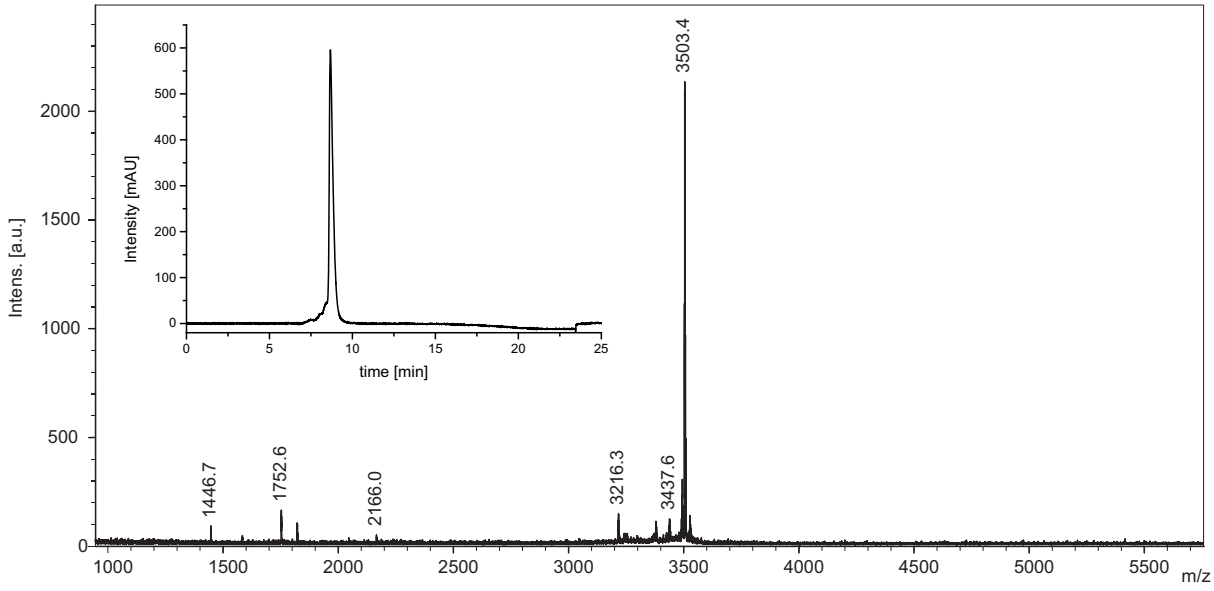


Figure A2: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA2-Cy5.

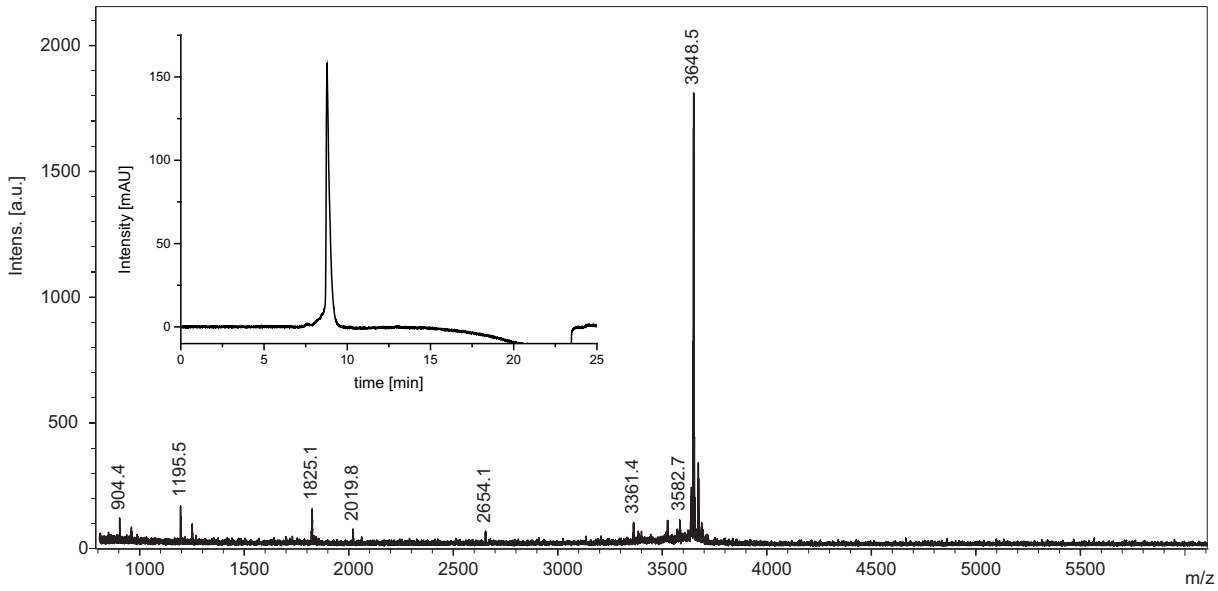


Figure A3: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA2S-Cy5.

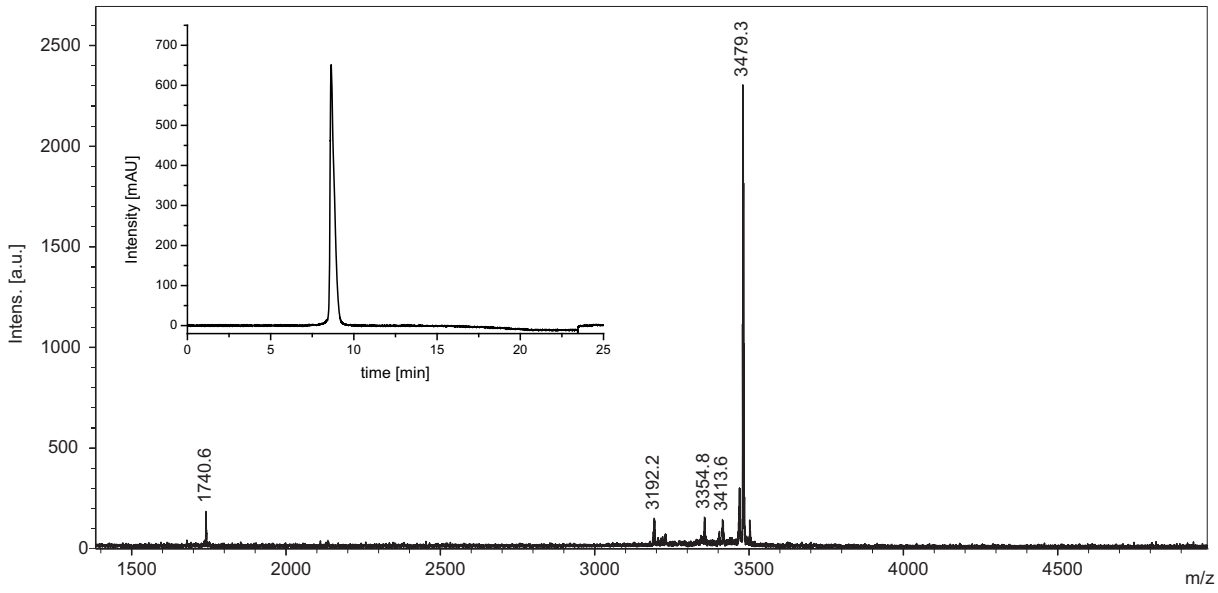


Figure A4: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA3-Cy5.

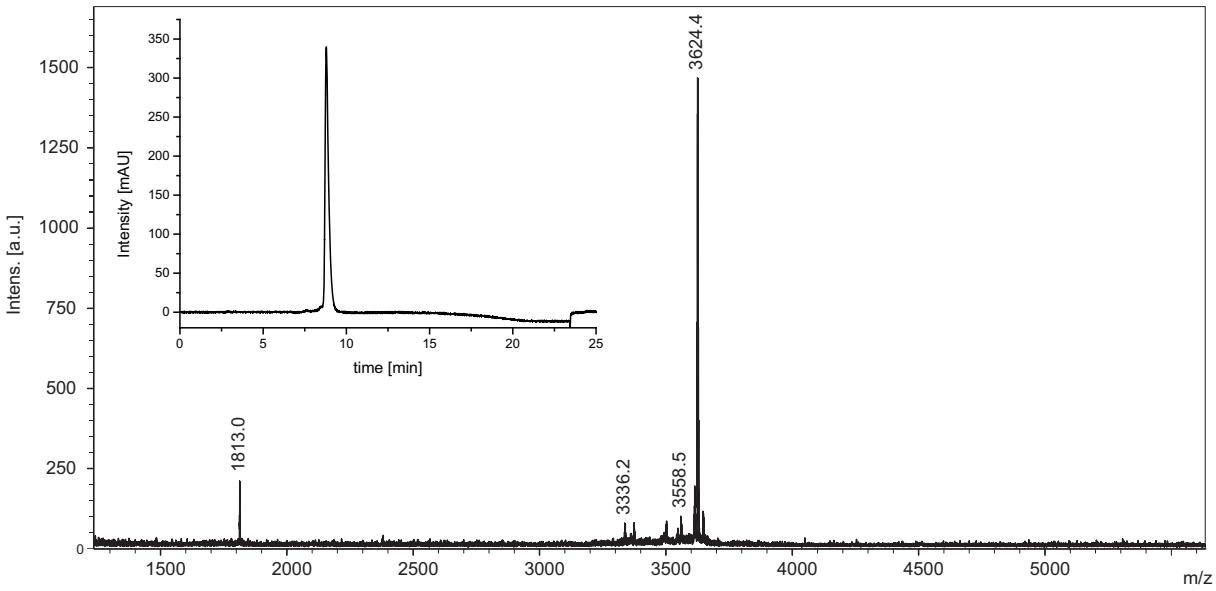


Figure A5: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA3S-Cy5.

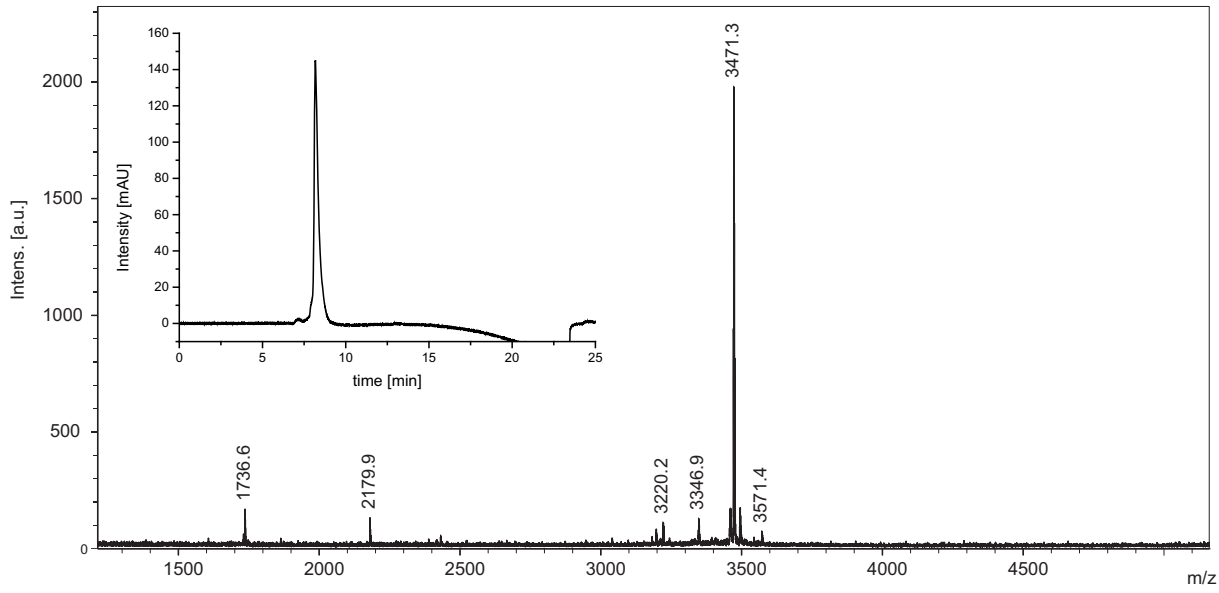


Figure A6: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA4-Cy3.

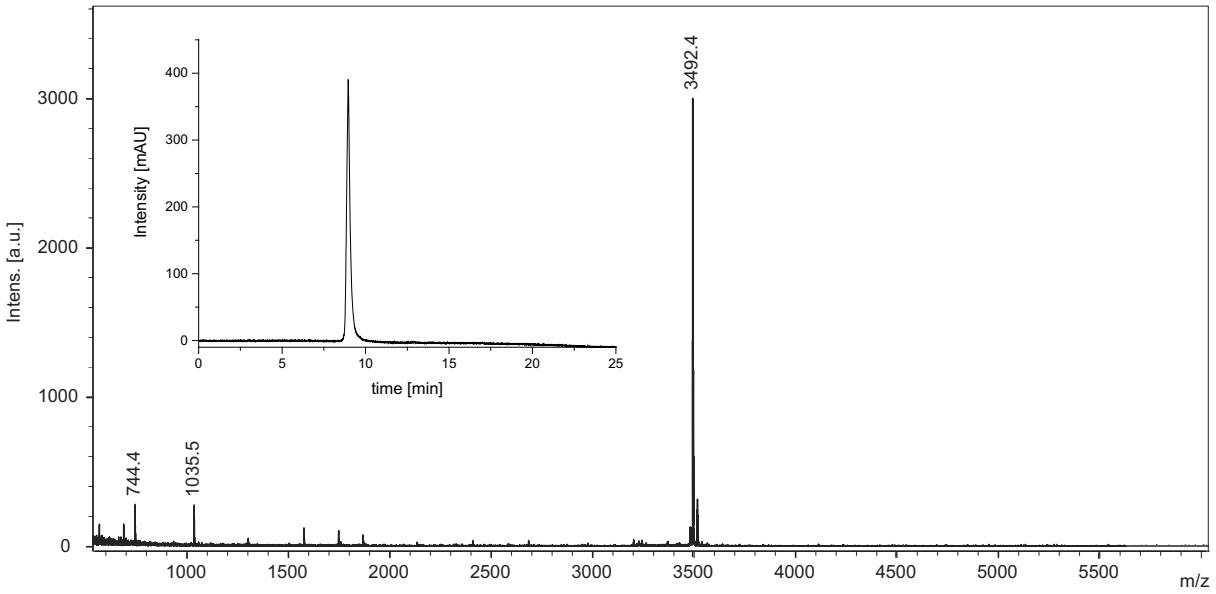


Figure A7: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA-Ex1-Cy3.

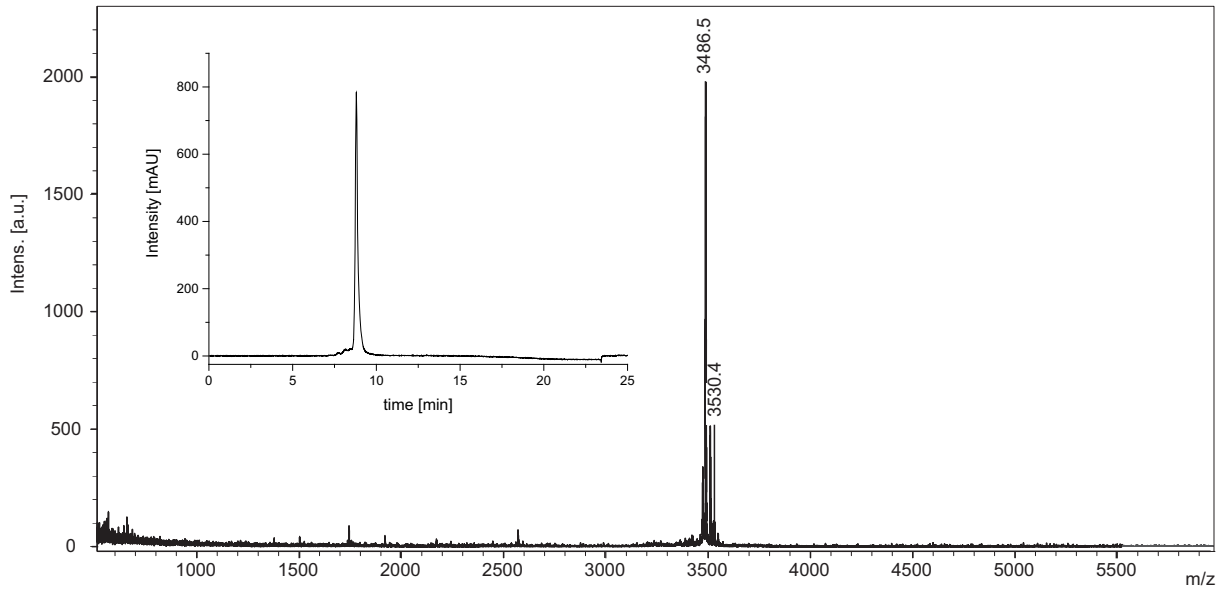


Figure A8: MALDI-MS spectrum and HPLC chromatogram (inset) of PNA-Ex2-Cy5.

B. Melting Curves

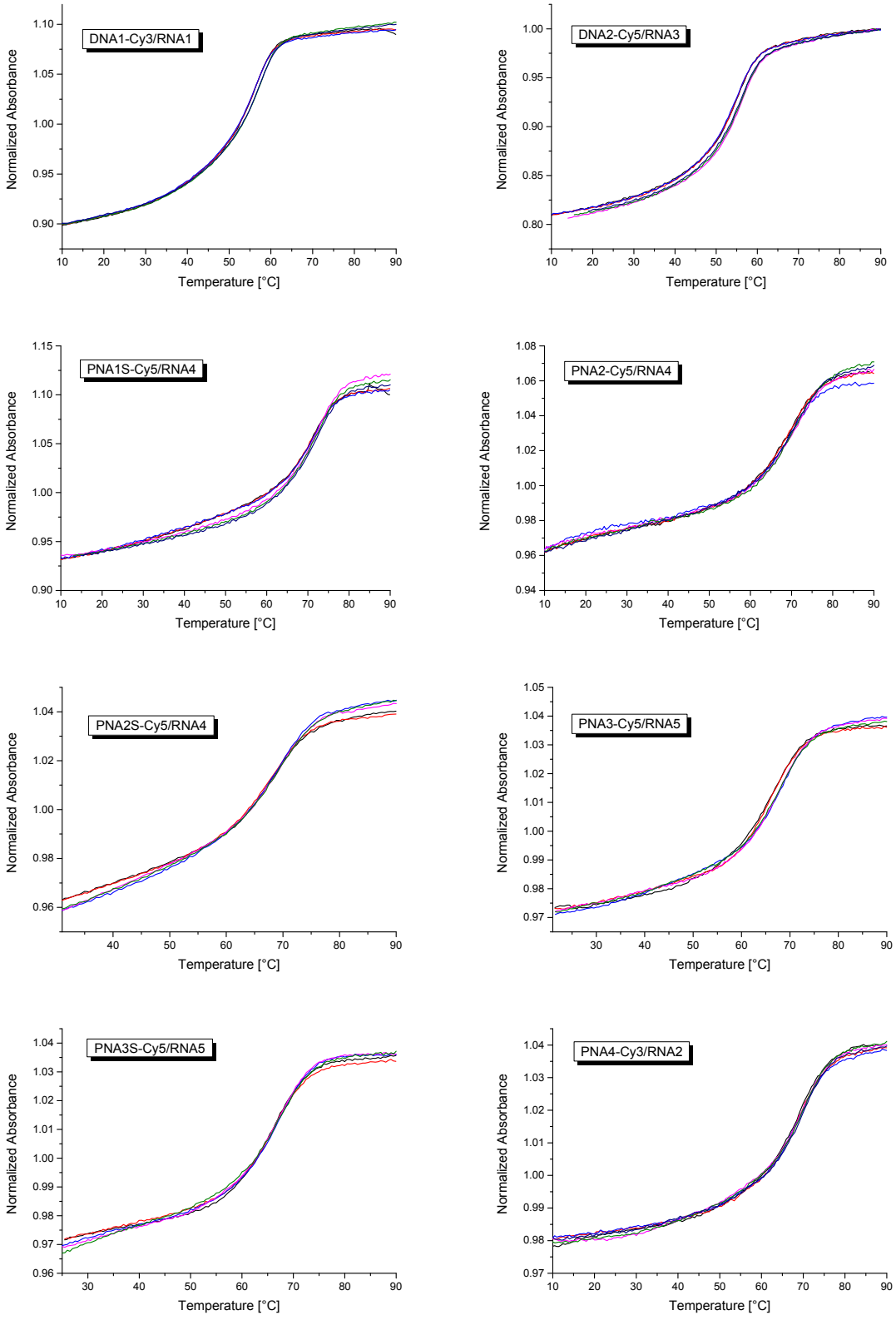


Figure A9: Melting curves of PNA-Cy and DNA-Cy derivatives with complementary RNA oligonucleotides recorded at 260 nm. Three cycles of heating and cooling. The absorbance was normalized (divided by mean). For exact conditions see Chapter 5.4.3.

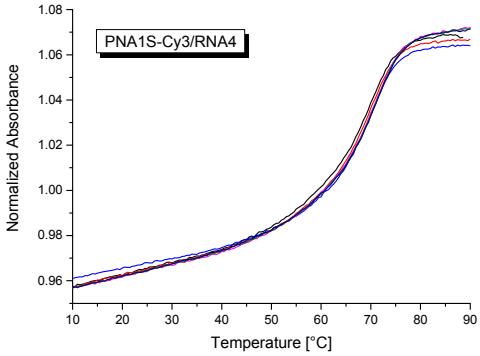


Figure A10: Melting curves of PNA1S-Cy3 with complementary RNA4 recorded at 260 nm. Three cycles of heating and cooling. The absorbance was normalized (divided by mean). For exact conditions see Chapter 5.4.3.

C. Single-Turnover Cleavage Assays

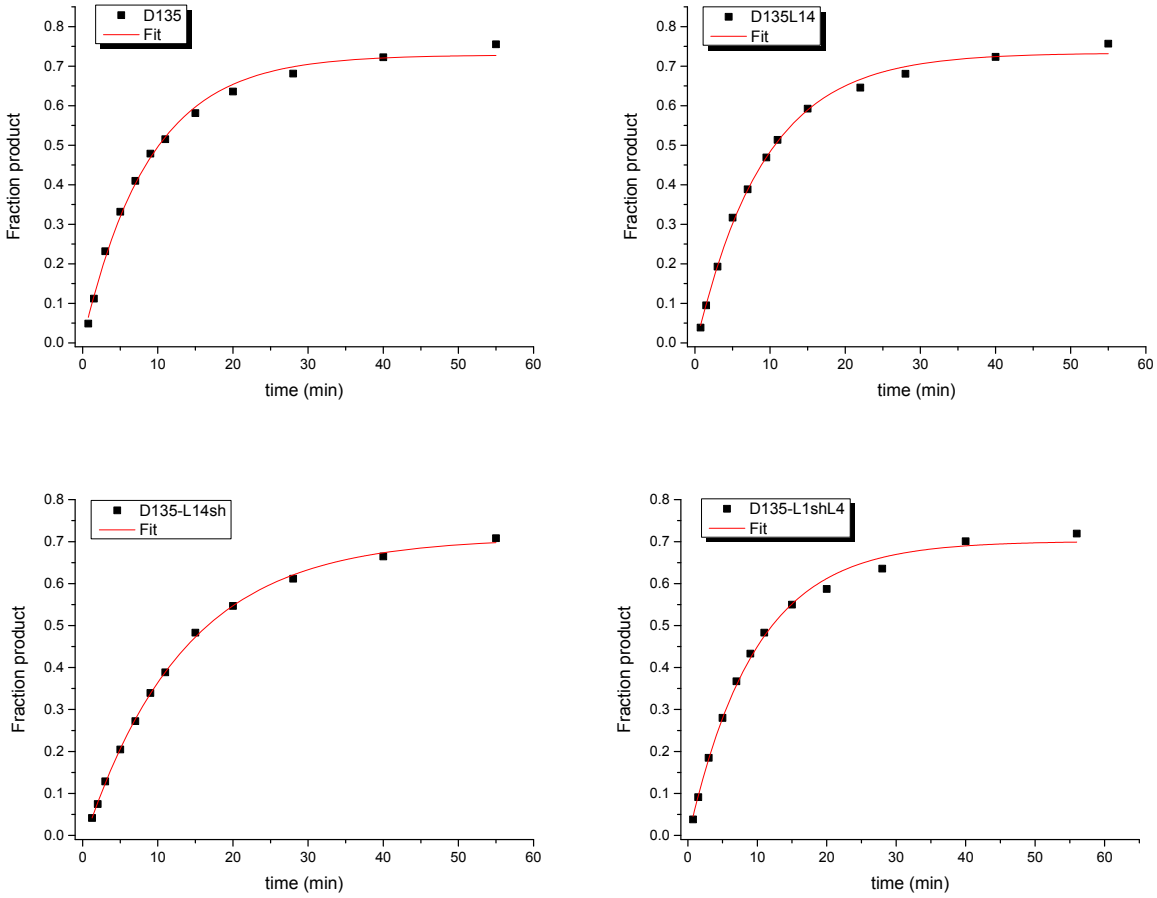


Figure A11: Cleavage of 17/7 substrate by different ribozyme constructs. Plots of product fraction versus time with exponential fit (red). For exact conditions see Chapter 5.4.6.

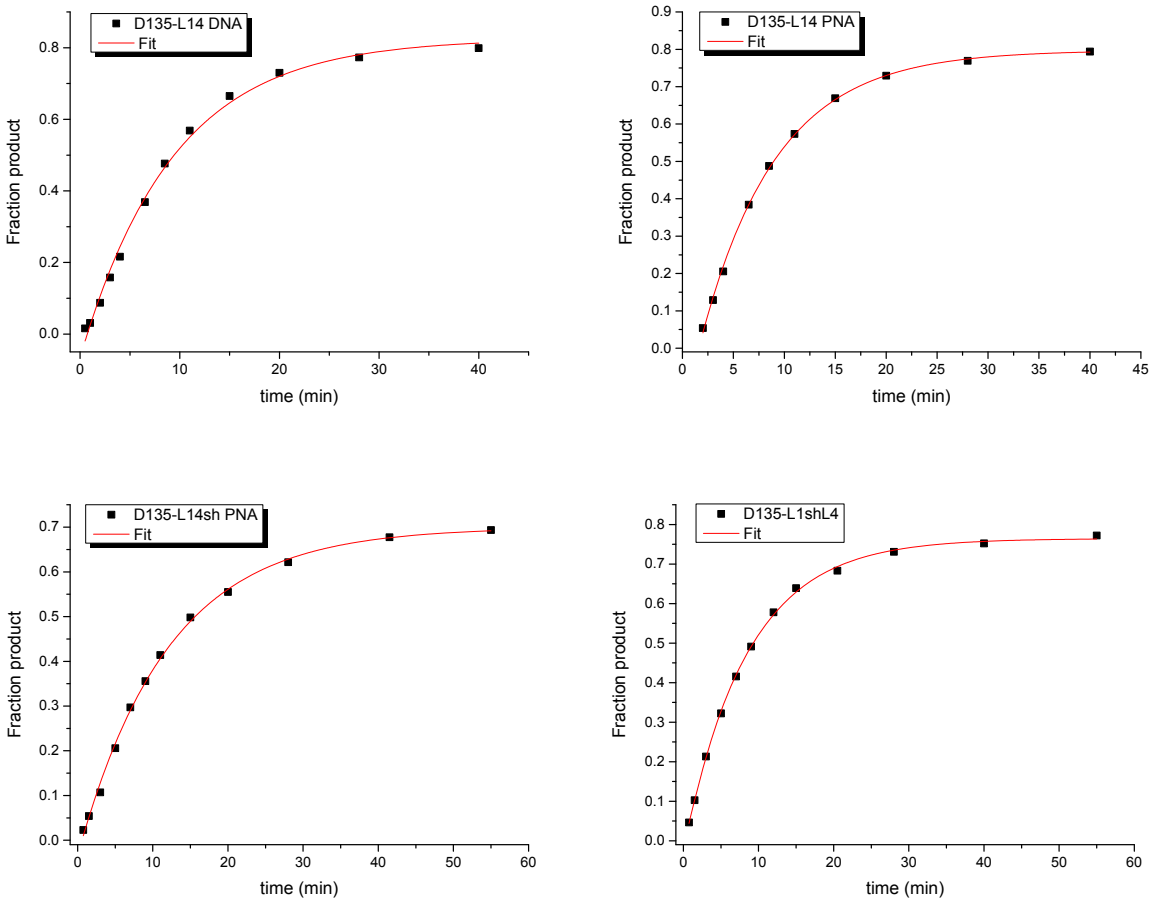


Figure A12: Cleavage of 17/7 substrate by different ribozyme constructs. Plots of product fraction versus time with exponential fit (red). For exact conditions see Chapter 5.4.6.

D. Plasmid Table pSPA7a

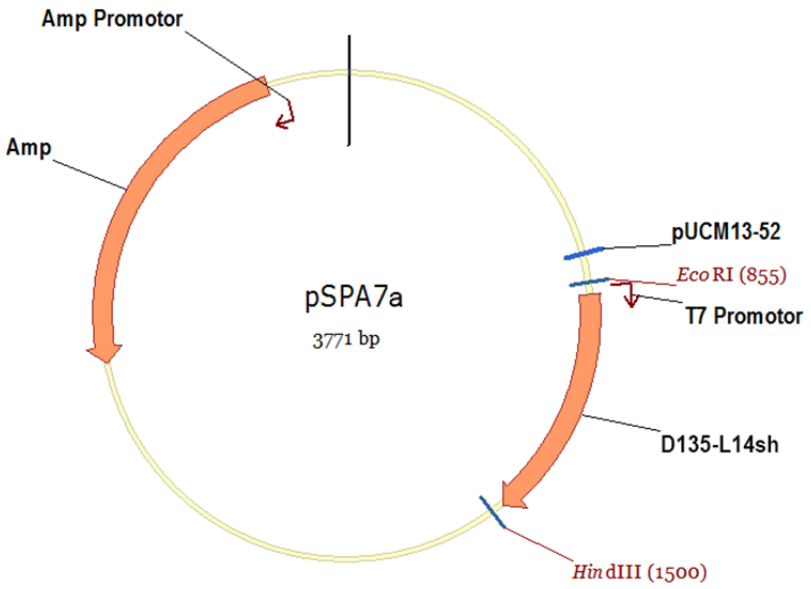


Figure A13: Plasmid pSPA7a. The plasmid pSP7a contains the ribozyme construct D135-L14sh.

1	TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG
	GAGACGGTCA	CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAGCCCG
101	TCAGGGCGCG	TCAGCGGGTG	TTGGCGGGTG	TCGGGGCTGG	CTTAACTATG
	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	ACCATATGCG	GTGTGAAATA
201	CCGCACAGAT	GCGTAAGGAG	AAAATACCGC	ATCAGGCGAC	GCGCCCTGTA
	GCGGCGCATT	AAGCGCGGCG	GGTGTGGTGG	TTACGCGCAG	CGTGACCGCT
301	ACACTTGCCA	GCGCCCTAGC	GCCCGCTCCT	TTCGCTTTCT	TCCCTTCCTT
	TCTCGCCACG	TTCGCCGGCT	TTCCCCGTCA	AGCTCTAAAT	CGGGGGCTCC
401	CTTTAGGGTT	CCGATTTAGT	GCTTTACGGC	ACCTCGACCC	CAAAAAACTT
	GATTAGGGTG	ATGGTTCACG	TAGTGGGCCA	TCGCCCTGAT	AGACGGTTTT
501	TCGCCCTTTG	ACGTTGGAGT	CCACGTTCTT	TAATAGTGGA	CTCTTGTTCC
	AAACTGGAAC	AACACTCAAC	CCTATCTCGG	TCTATTCTTT	TGATTTATAA
601	GGGATTTTGC	CGATTTCGGC	CTATTGGTTA	AAAAATGAGC	TGATTTAACA
	AAAATTTAAC	GCGAATTTTA	ACAAAATATT	AACGCTTACA	ATTTCCATTC
701	GCCATTCAGG	CTGCGCAACT	GTTGGGAAGG	GCGATCGGTG	CGGGCCTCTT
	CGCTATTACG	CCAGCTGGCG	AAAGGGGGAT	GTGCTGCAAG	GCGATTAAGT
801	TGGGTAACGC	CAGGGTTTTC	CCAGTCACGA	CGTTGTAAAA	CGACGGCCAG
	EcoRI			D135-L14sh	
	GGCGAATTCG	TGAATTGTAA	TACGACTCAC	TATAGAGCGG	TCTGAAAGTT
901	ATCATAAATA	ATATTTACCA	TATAATAATG	GATAAATTAT	ATTTTTATCA
	ATATAAGTCT	AATTACAAGT	GTATTAAAAT	GGTAACATAA	ATATGCTAAG
1001	CTGTAATGAC	AAAAGTATCC	ATATTCTTGA	CAGTTATTTT	ATATTATAAA
	AAAAAGATGA	AGGAACTTTG	ACTGATCTAA	TATGCTCAAC	GAAAGTGAAT

```
1101  CAAATGTTAT  AAAATTACTT  ACACCACTAA  TTGAAAACCT  GTCTGATATT
      CAATTATTAT  CCCACGCTCT  TGGTAGGGAT  AAAATGGTTG  ATGTTATGTA
1201  TTGGAAATGA  GCATACGATA  AATCATATAA  CCATTAGTAA  TATAATTTGA
      GAGCTAAGTT  AGATATTTAC  GTATTTATGA  TAAAACAGAA  TAAACCCTCT
1301  TCGGAGGGTA  AAAGATTGTA  TAAAAAGCTA  ATGCCATATT  GTAATGATAT
      GGATAAGAAT  TATTATTCTA  AAGATGAAAA  TCTGCTAACT  TATACTATAG
1401  GTGATACCCA  CGGACTACGT  AGGGTATTGC  GTGAGCCGTA  TGCGATGAAA
                                                    HindIII
      GTCGCACGTA  CGGTTCCTGG  ATCCTCTAGA  GTCGACCTGC  AGGCATGCAA
1501  GCTTTTGTTC  CCTTTAGTGA  GGGTTAATTC  CGAGCTTGGC  GTAATCATGG
      TCATAGCTGT  TTCCTGTGTG  AAATTGTTAT  CCGCTCACAA  TTCCACACAA
1601  CATACGAGCC  GGAAGCATAA  AGTGTAAAGC  CTGGGGTGCC  TAATGAGTGA
      GCTAACTCAC  ATTAATTGCG  TTGCGCTCAC  TGCCCGCTTT  CCAGTCGGGA
1701  AACCTGTCGT  GCCAGCTGCA  TTAATGAATC  GGCCAACGCG  CGGGGAGAGG
      CGGTTTGCGT  ATTGGGCGCT  CTTCCGCTTC  CTCGCTCACT  GACTCGCTGC
1801  GCTCGGTCGT  TCGGCTGCGG  CGAGCGGTAT  CAGCTCACTC  AAAGGCGGTA
      ATACGGTTAT  CCACAGAATC  AGGGGATAAC  GCAGGAAAGA  ACATGTGAGC
1901  AAAAGGCCAG  CAAAAGGCCA  GGAACCGTAA  AAAGGCCGCG  TTGCTGGCGT
      TTTTCCATAG  GCTCCGCCCC  CCTGACGAGC  ATCACAAAAA  TCGACGCTCA
2001  AGTCAGAGGT  GGCGAAACCC  GACAGGACTA  TAAAGATACC  AGGCGTTTCC
      CCCTGGAAGC  TCCCTCGTGC  GCTCTCCTGT  TCCGACCCTG  CCGCTTACCG
2101  GATACCTGTC  CGCCTTTCTC  CCTTCGGGAA  GCGTGGCGCT  TTCTCATAGC
      TCACGCTGTA  GGTATCTCAG  TTCGGTGTAG  GTCGTTCGCT  CCAAGCTGGG
2201  CTGTGTGCAC  GAACCCCCCG  TTCAGCCCGA  CCGCTGCGCC  TTATCCGGTA
      ACTATCGTCT  TGAGTCCAAC  CCGGTAAGAC  ACGACTTATC  GCCACTGGCA
2301  GCAGCCACTG  GTAACAGGAT  TAGCAGAGCG  AGGTATGTAG  GCGGTGCTAC
      AGAGTTCTTG  AAGTGGTGGC  CTAACTACGG  CTACACTAGA  AGGACAGTAT
2401  TTGGTATCTG  CGCTCTGCTG  AAGCCAGTTA  CCTTCGGAAA  AAGAGTTGGT
      AGCTCTTGAT  CCGGCAAACA  AACCACCGCT  GGTAGCGGTG  GTTTTTTTGT
2501  TTGCAAGCAG  CAGATTACGC  GCAGAAAAAA  AGGATCTCAA  GAAGATCCTT
      TGATCTTTTC  TACGGGGTCT  GACGCTCAGT  GGAACGAAAA  CTCACGTTAA
2601  GGGATTTTGG  TCATGAGATT  ATCAAAAAGG  ATCTTCACCT  AGATCCTTTT
      AAATTAAAAA  TGAAGTTTTA  AATCAATCTA  AAGTATATAT  GAGTAAACTT
2701  GGTCTGACAG  TTACCAATGC  TTAATCAGTG  AGGCACCTAT  CTCAGCGATC
      TGTCTATTTC  GTTCATCCAT  AGTTGCCTGA  CTCCCCGTCG  TGTAGATAAC
2801  TACGATACGG  GAGGGCTTAC  CATCTGGCCC  CAGTGCTGCA  ATGATACCGC
      GAGACCCACG  CTCACCGGCT  CCAGATTTAT  CAGCAATAAA  CCAGCCAGCC
2901  GGAAGGGCCG  AGCGCAGAAG  TGGTCCTGCA  ACTTTATCCG  CCTCCATCCA
      GTCTATTAAT  TGTTGCCGGG  AAGCTAGAGT  AAGTAGTTCG  CCAGTTAATA
3001  GTTTGCGCAA  CGTTGTTGCC  ATTGCTACAG  GCATCGTGGT  GTCACGCTCG
```

```
      TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT
3101  TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC
      CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG
3201  GCAGCACTGC ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC
      TGTGACTGGT GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC
3301  GACCGAGTTG CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT
      AGCAGAACTT TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA
3401  ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC
      GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG
3501  TGAGCAAAAA CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC
      ACGGAAATGT TGAATACTCA TACTCTTCCT TTTTCAATAT TATTGAAGCA
3601  TTTATCAGGG TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG
      AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA AAGTGCCACC
3701  TGACGTCTAA GAAACCATTA TTATCATGAC ATTAACCTAT AAAAATAGGC
      GTATCACGAG GCCCTTTCGT C
```

The plasmid pSPA7b is analogous to pSPA7a encoding D135-L1shL4 instead of D135-L14sh.

E. UV/Vis Spectra

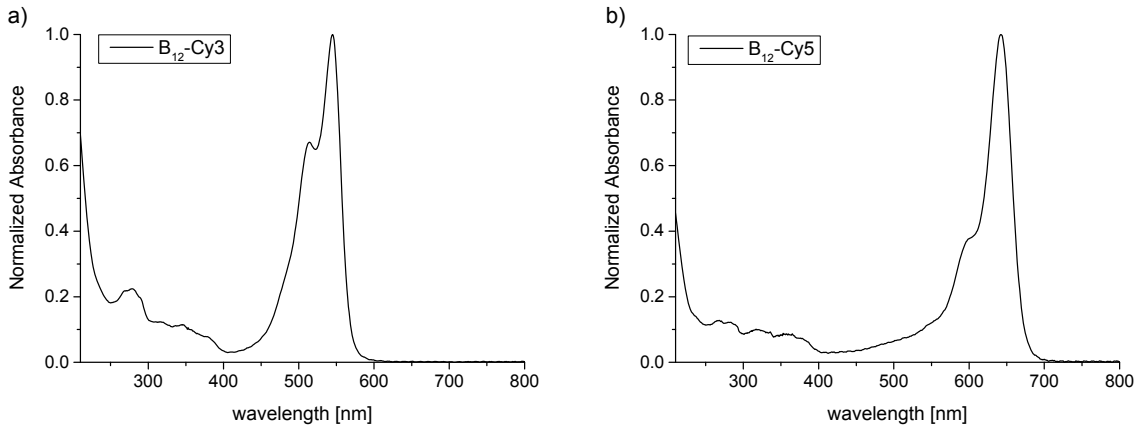


Figure A14: UV/Vis spectra of B_{12}-Cy3 and B_{12}-Cy5. Normalized absorbance spectra of a) **B_{12}-Cy3** and b) **B_{12}-Cy5** in water at r.t.

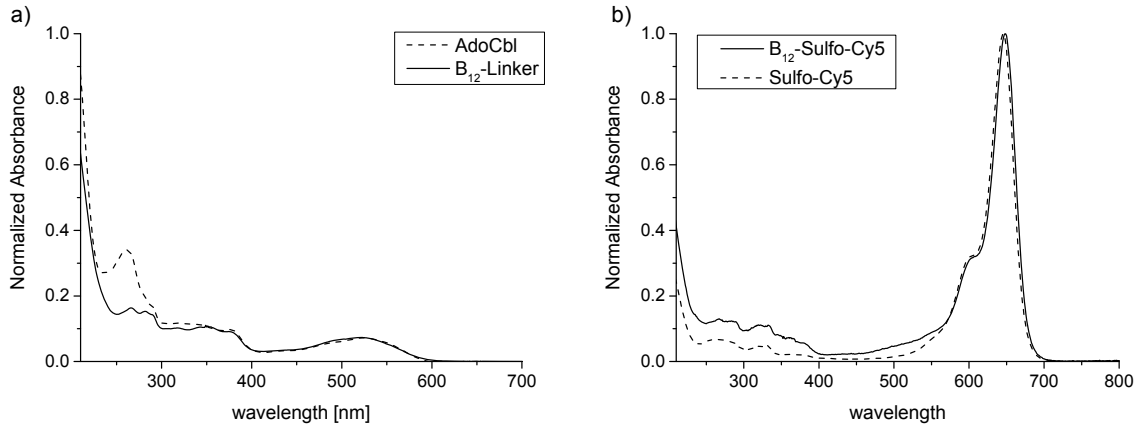


Figure A15: UV/Vis spectra of B_{12} derivatives. Normalized absorbance spectra of a) AdoCbl (dotted line) and **B_{12}-linker** (solid line), and b) Sulfo-Cy5 (dotted line) and **B_{12}-Sulfo-Cy5** (solid line) in water at r.t.

F. HR-ESI-MS Spectra

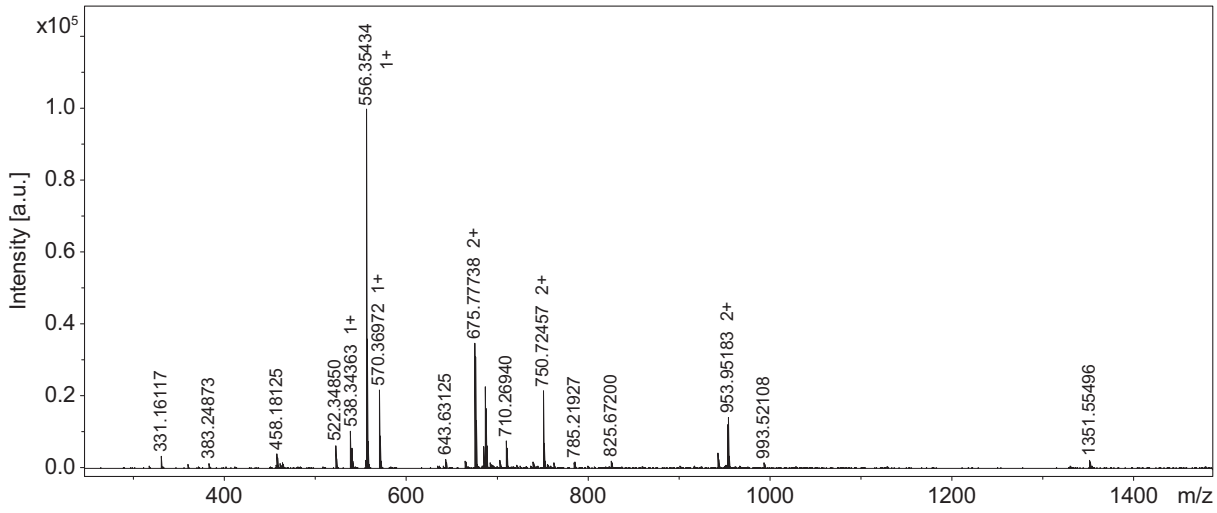


Figure A16: HR-ESI-MS spectrum of B_{12}-Cy5.

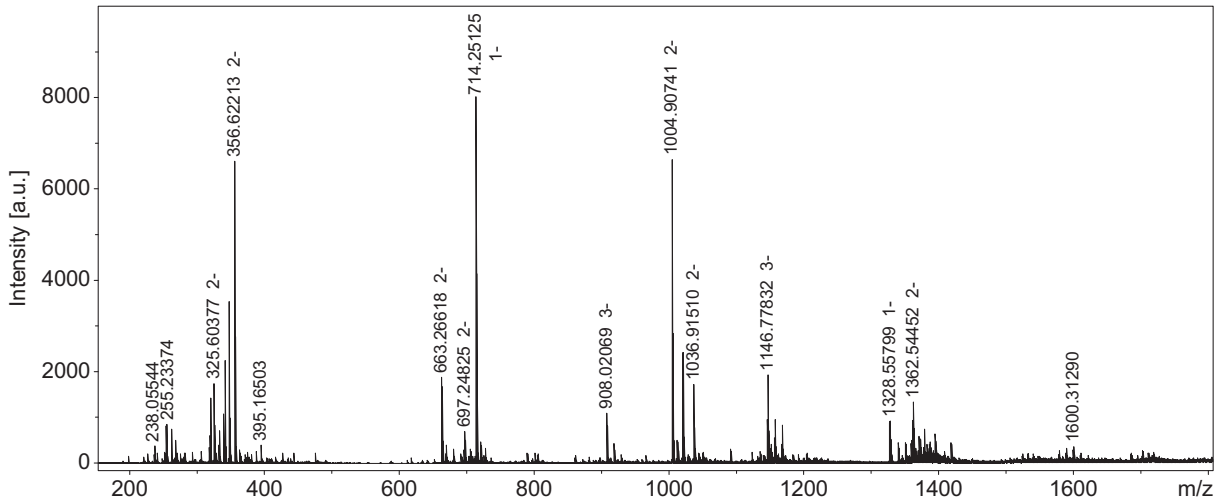


Figure A17: HR-ESI-MS spectrum of B_{12}-Sulfo-Cy5.

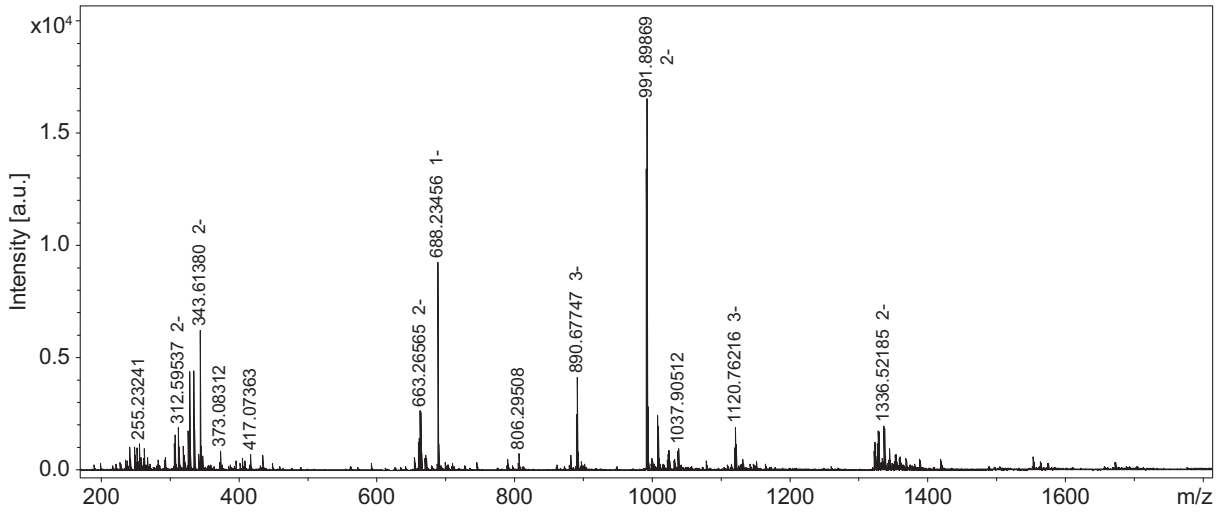


Figure A18: HR-ESI-MS spectrum of B_{12}-Sulfo-Cy3.

G. NMR Spectra

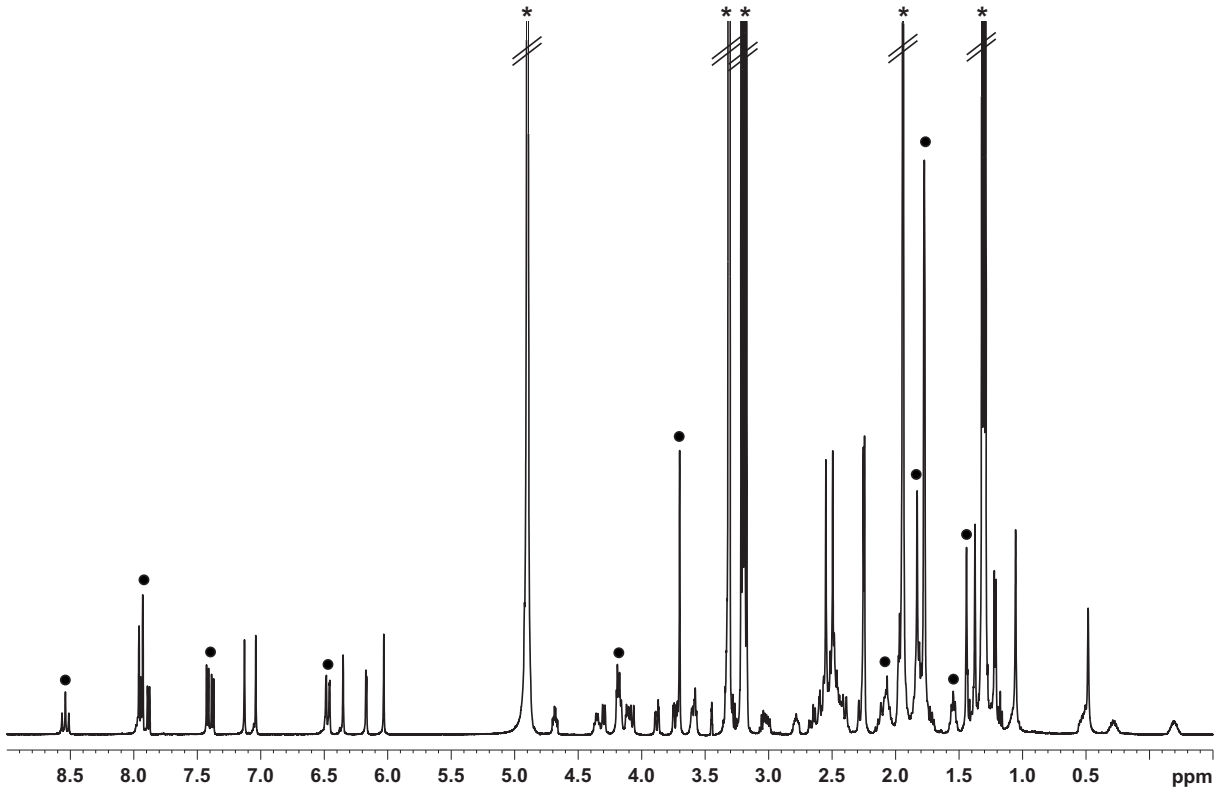


Figure A19: ^{1}H-NMR spectrum of B_{12}-Sulfo-Cy3 in MeOD-d_4 (500 MHz, 300 K). Protons belonging to Sulfo-Cy3 are marked with black dots (●).

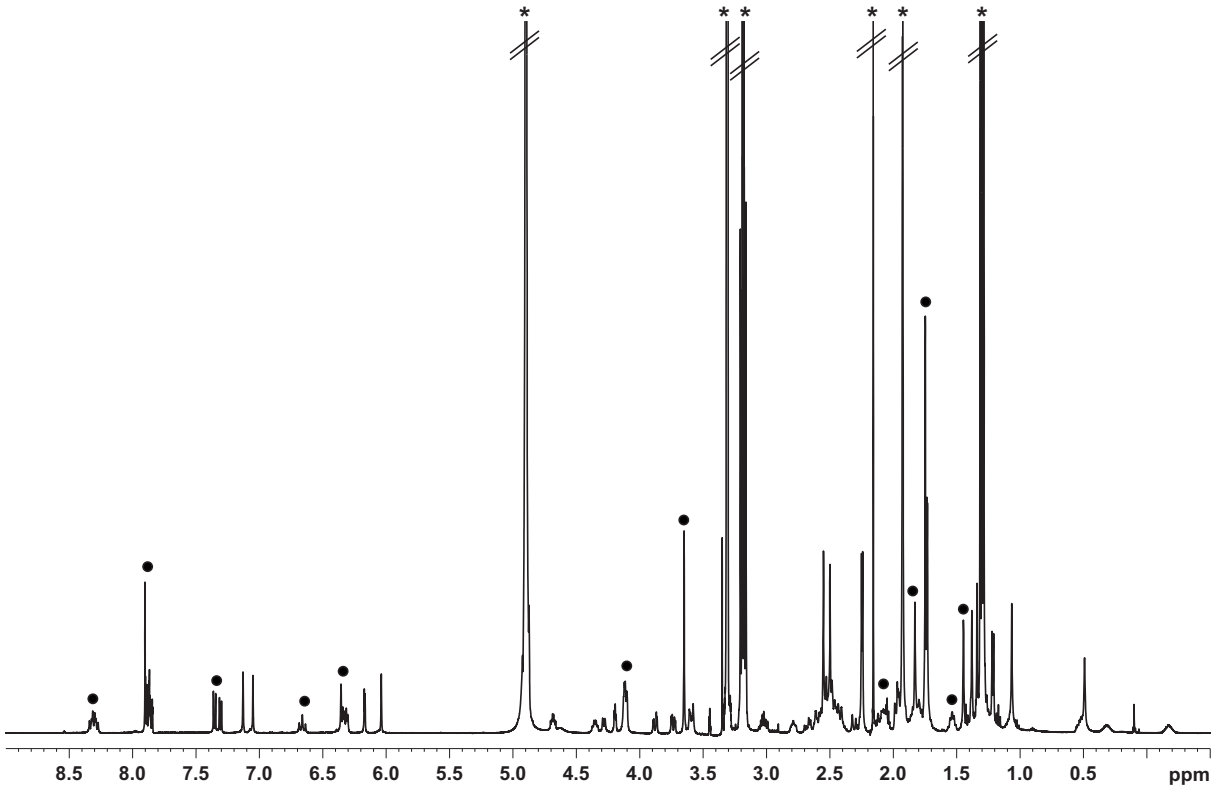


Figure A20: ^{1}H-NMR spectrum of B_{12}-Sulfo-Cy5 in MeOD-d_4 (500 MHz, 300 K). Protons belonging to Sulfo-Cy5 are marked with black dots (●).

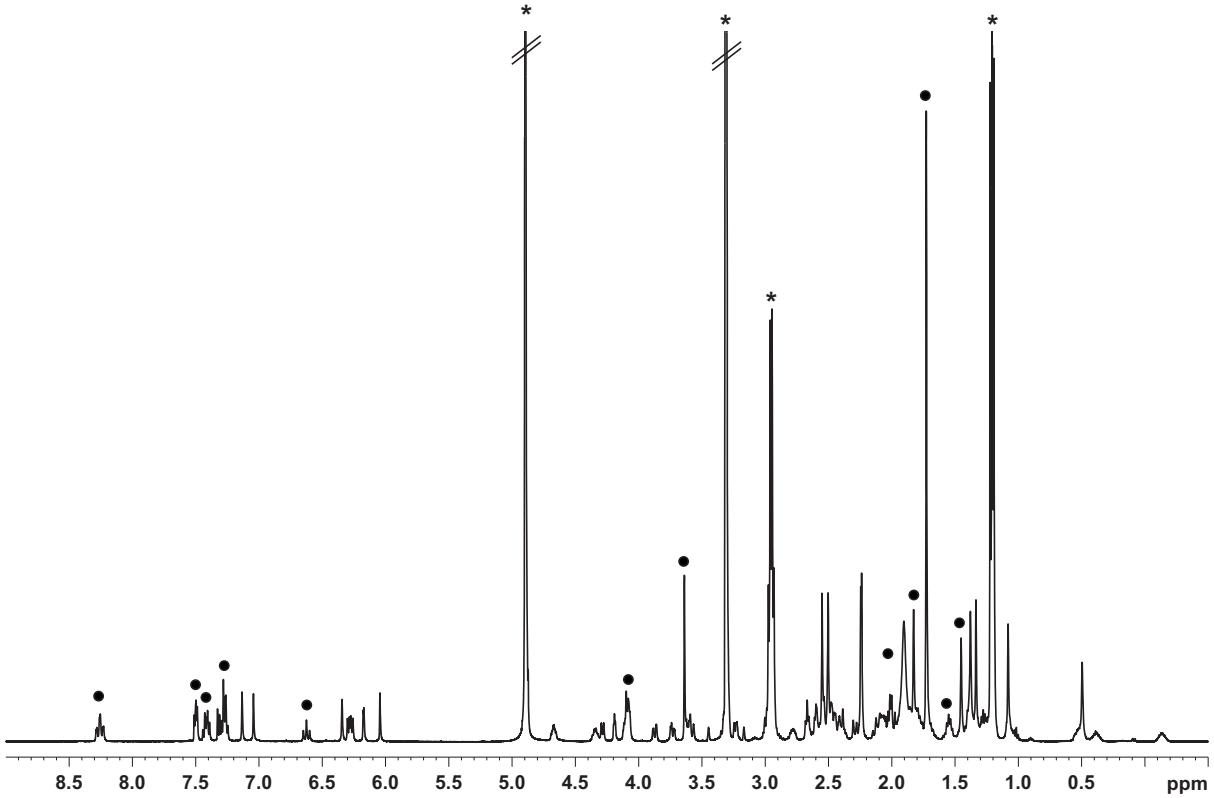


Figure A21: ^{1}H-NMR spectrum of B_{12}-Cy5 in MeOD-d_4 (500 MHz, 300 K). Protons belonging to Cy5 are highlighted with black dots (●).

H. Time Trajectories of smFRET Experiments

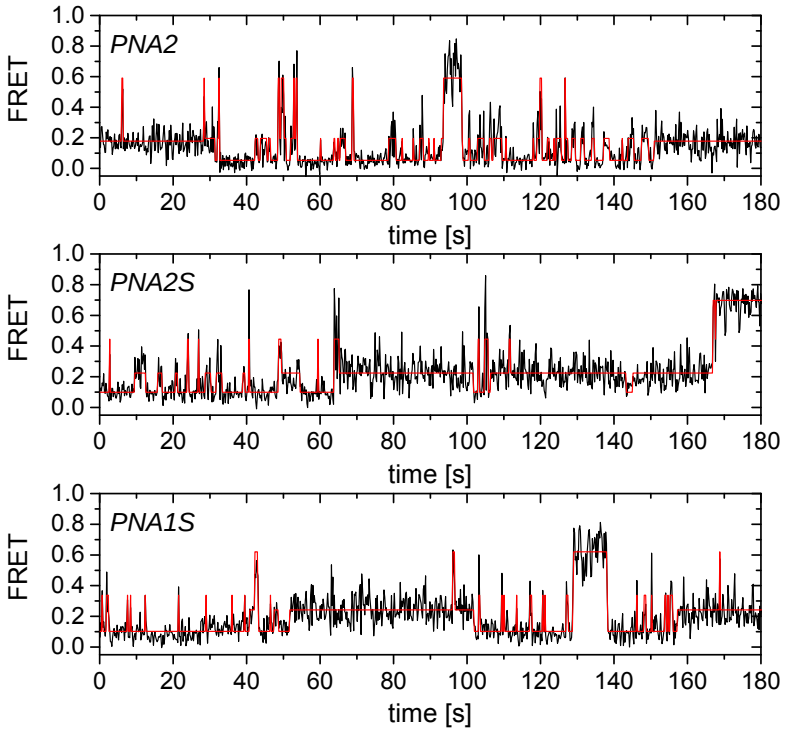


Figure A22: Characteristic time trajectories of smFRET experiments of D135-L14 labeled with PNA2-Cy5, PNA2S-Cy5, or PNA1S-Cy5. The different conformational states at 0.1, 0.2, 0.4 and 0.6 are determined by vbFRET software based on HMM (red line).

I. btuB Constructs

pPC1

The *btuB* riboswitch (*E. coli*) construct pPC1 (Figure A23) encoded in the plasmid pPC1 is designed by *Choudhary et al*.[176] To improve the transcription start site for the T7 RNA polymerase, the construct contains an insertion of three nucleotides (GGA) between the TATA box and the *btuB* coding sequence. Furthermore, the plasmid contains the *btuB* aptamer, and *SacI* and *EcoR1* restriction sites located downstream of the riboswitch region.

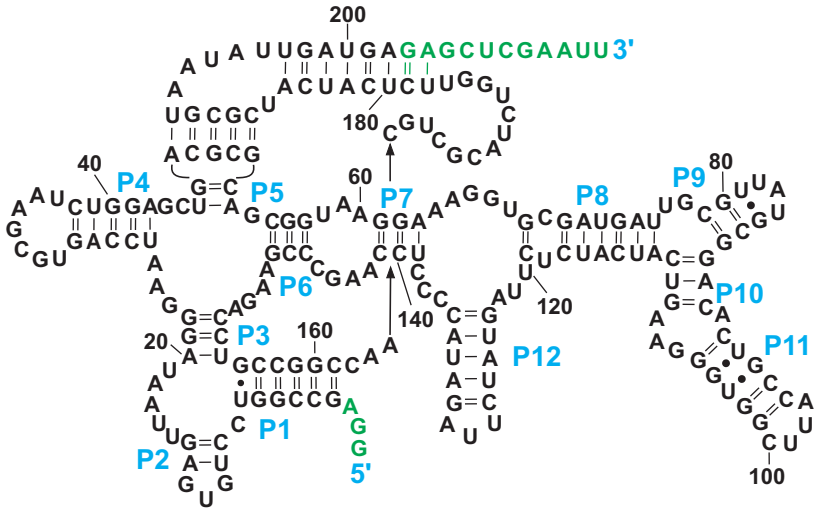


Figure A23: Secondary structure of the *btuB* riboswitch (*E. coli*) construct pPC1. The *btuB* aptamer is depicted in black, extensions are in green and characteristic structure elements are in blue. The numbering corresponds to the wild-type sequence. 216 nts

pPC4

The *btuB* riboswitch (*E. coli*) construct pPC4 (Figure A24) encoded in the plasmid pPC4 is designed by *Choudhary et al*.[176] To improve the transcription start site for the T7 RNA polymerase, the construct contains an insertion of three nucleotides (GGA) between the TATA box and the *btuB* coding sequence. Furthermore, an artificial loop for binding of oligonucleotides for labeling purpose is inserted. The plasmid contains the modified *btuB* aptamer and expression platform, and an *EcoR1* restriction site located downstream of the riboswitch region.

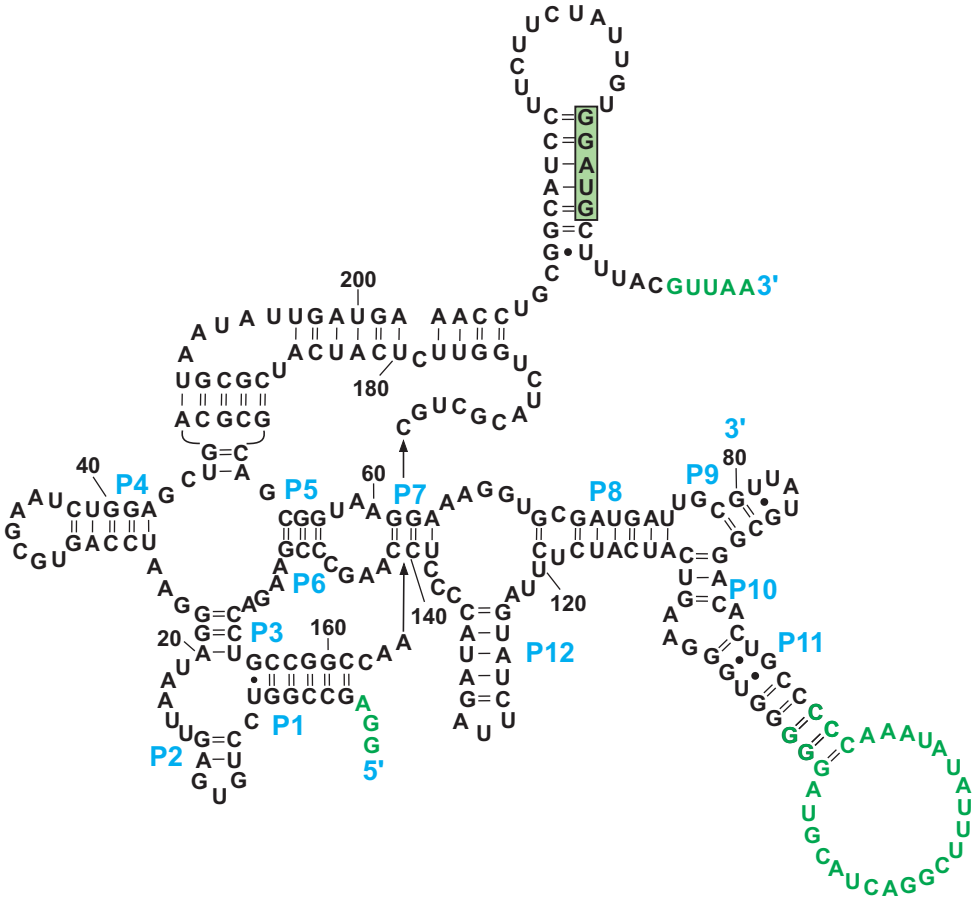


Figure A24: Secondary structure of the *btuB* riboswitch (*E. coli*) construct pPC4. The wild-type sequence of the *btuB* riboswitch is depicted in black, insertions and extensions are in green, characteristic structure elements are in blue and the ribosomal binding site is highlighted in the green box. The numbering corresponds to the wild-type sequence. 270 nts

pSG1

The *btuB* riboswitch (*E. coli*) construct pSG1 (Figure A25) encoded in the plasmid pSG1 is designed by Gallo *et al.*[87] The plasmid contains the *btuB* aptamer, the self-cleaving hammerhead ribozyme sequence at the 5'- and the self-cleaving hepatitis delta virus ribozyme sequence at the 3'-site, which lead to uniform RNA ends after transcription and cleavage, and additionally an *EcoR1* restriction site located downstream of the riboswitch region.

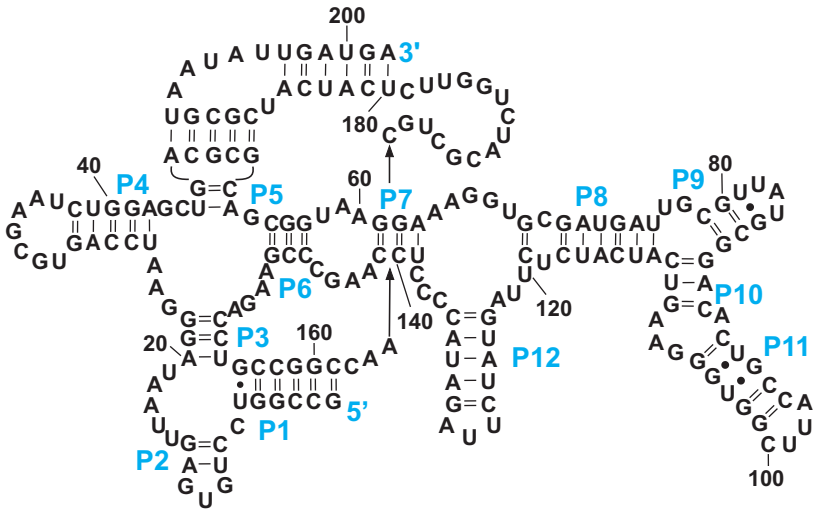


Figure A25: Secondary structure of the *btuB* riboswitch (*E. coli*) construct pSG1. The *btuB* aptamer is depicted in black and characteristic structure elements are in blue. 202 nts. *Figure adapted from Dr. Sofia Gallo.*[87]

10 Curriculum Vitae

Personal Data

Name	Anita G. Schmitz
Date and place of birth	16th December 1984 in Erlangen, Germany
Nationality	German

Education

07/2010 - 09/2014	PhD studies (University of Zurich, Switzerland) in the group of Prof. Roland K. O. Sigel at the Department of Chemistry, on "*New Labeling Strategies to Investigate the Folding of Large RNAs via FRET*"
10/2004 – 09/2009	Diploma studies in chemistry (Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Germany) Diploma thesis in the group of Prof. Ivana Ivanović-Burmazović on "*Cobalt(II)- und Cobalt(III)-Porphyrins: Synthesis, Characterization und Reaction with Superoxide*"
1995 – 2004	Secondary school (Gymnasium Eckental) Abitur
1991 – 1995	Elementary school (Volksschule Eckental–Eckenhaid)

Awards and Scholarships

2013	Poster prize at the 8th Annual Dorothy Crowfoot Hodgkin Symposium (Zurich, Switzerland)
2011	Poster prize at the CMSZH Graduate School Retreat (Randa, Switzerland)
2009	Member of Leonardo-Kolleg University Erlangen-Nürnberg (exclusive for top 5% of all students)

Publications

"Internal Labeling Strategy of Large RNAs with Minimal Perturbation Using Fluorescent PNA"
Anita G. Schmitz, Susann Zelger-Paulus, Gilles Gasser, Roland K. O. Sigel, *ChemBioChem* **2015**, accepted.

"Photo-induced uncaging of a specific Re(I) organometallic complex in living cells"
Anna Leonidova, Vanessa Pierroz, Riccardo Rubbiani, Yanjun Lan, Anita G. Schmitz, Andres Kaech, Roland K. O. Sigel, Stefano Ferrari, Gilles Gasser, *Chemical Science* **2014**, *5*, 4044-4056.

Conference Contributions

Fluorescent Labels for Single Molecule Studies of RNA Structure and Folding
Anita G. Schmitz, Susann Zelger-Paulus, Gilles Gasser, Roland K. O. Sigel (*poster*)
Fall Meeting of the Swiss Chemical Society, Zurich, Switzerland, 2014.

Peptide Nucleic Acids - Superior FRET-Labels for Single Molecule Studies of RNA Structure and Folding
Anita G. Schmitz, Susann Zelger-Paulus, Philipp Anstaett, Gilles Gasser, Roland K. O. Sigel (*poster*)
12th European Biological Inorganic Chemistry Conference (EUROBIC), Zurich, Switzerland, 2014.

Elucidation of the btuB Riboswitch Folding with Fluorescent B_{12} via FRET
Anita G. Schmitz, Susann Zelger-Paulus, Philipp Anstaett, Gilles Gasser, Roland K. O. Sigel (*poster*)
2nd International Symposium on Functional Metal Complexes that Bind to Biomolecules (3rd Whole Action Meeting of the COST Action CM1105), Zurich, Switzerland, 2014.

Unnatural smFRET Labels for Elucidating the Natural RNA Folds - An Unconventional Approach
Anita G. Schmitz, Susann Paulus, Gilles Gasser, Roland K. O. Sigel (*oral presentation*)
12th International Symposium on Applied Bioinorganic Chemistry (ISABC), Guangzhou, China, 2013.

Specific RNA Labeling Strategies for Fluorescence Applications
Anita G. Schmitz, Susann Paulus, Gilles Gasser, Roland K. O. Sigel (*poster*)
8th Annual Dorothy Crowfoot Hodgkin Symposium, Zürich, Switzerland, 2013.

Specific RNA Labeling Strategies for Fluorescence Applications
Anita G. Schmitz, Gilles Gasser, Roland K. O. Sigel (*poster*)
Fall Meeting of the Swiss Chemical Society, Lausanne, Switzerland, 2013.

New labeling strategies to investigate the B_{12}-binding to the btuB riboswitch of E. coli via FRET
Anita Schmitz, Gilles Gasser, Roland K. O. Sigel (*oral presentation*)
Retreat of the CMSZH Graduate School, Charmey, Switzerland, 2013.

The Sigel lab – Metal ions and a bit of chemistry

Erica Fiorini, Helena Guiset Micherachs, Melodie Hadzic, Mokrane Khier, Sebastian König, Anastasia Musiari, Susann Paulus, Anita Schmitz, Marianthi Zampakou, Roland K. O. Sigel (*poster*)
Retreat of the CMSZH Graduate School, Charmey, Switzerland, 2013.

New labeling strategies to investigate the B_{12}-binding to the btuB riboswitch of E. coli via FRET
Anita Schmitz, Gilles Gasser, Roland K. O. Sigel (*poster*)
7th Annual Dorothy Crowfoot Hodgkin Symposium, Zürich, Switzerland, 2012.

New labeling strategies to investigate the B_{12}-binding to the btuB riboswitch of E. coli via FRET
Anita Schmitz, Gilles Gasser, Roland K. O. Sigel (*poster*)
11th European Biological Inorganic Chemistry Conference (EUROBIC), Granada, Spain, 2012.

Investigations on the Interaction between the btuB Riboswitch and Fluorescent B_{12} Derivatives
Anita Schmitz, Roland K. O. Sigel (*poster*)
6th Annual Dorothy Crowfoot Hodgkin Symposium, Zürich, Switzerland, 2011.

Investigations on the Interaction between the btuB Riboswitch and Fluorescent B_{12} Derivatives
Anita Schmitz, Roland K. O. Sigel (*poster*)
Fall Meeting of the Swiss Chemical Society, Lausanne, Switzerland, 2011.

Investigation of B_{12}-binding to the btuB Riboswitch of E. coli via FRET
Anita Schmitz, Roland K. O. Sigel (*poster*)
15th International Conference on Biological Inorganic Chemistry (ICBIC), Vancouver, Canada, 2011.

The Sigel Group – Metallo RNA with a little bit of chemistry
Maximiliane Korth, Anita Schmitz, Sebastian König, Joachim Schnabl, Roland K. O. Sigel (*poster*)
Retreat of the CMSZH Graduate School, Randa, Switzerland, 2011.

Cobalt Porphyrins as Activators of Superoxide
Anita Schmitz, Nina Lang, Norbert Jux, Frank W. Heinemann, Ivana Ivanović-Burmazović (*poster*)
Inorganic/Bioinorganic Reaction Mechansims Group (IRMG) Meeting, Kloster Banz, Germany, 2010.

If you still wonder WHAT IS FRET …

… For sure it is more than just noise.